STIC Search Report
Biotech-Chem Library

Scientific & Technical Information Center
STIC
USPTO

TO: David Lamberston
Location: rem/2b79/2c70
Art Unit: 1636
Friday, April 01, 2005

Case Serial Number: 10/042059

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen 1a69
Phone: 571-272-2518

barbara.obryen@uspto.gov

Search Notes

Sira
Search and Information
Resources Administration

STIC-Biotech/ChemLib

148 695

From:	Lambertson, David
Sent:	Wednesday, March 23, 2005 1:32 PM
To:	STIC-Biotech/ChemLib
Cc:	Lambertson, David
Subject:	Search Request

Search Request

Examiner's Name:	David Lambertson
Examiner #:	79514
Art Unit:	1636
Room #:	Remsen 02B79
Mailbox room#:	Remsen 02C70
Phone:	(571) 272-0771
Results Format:	paper

Serial # 10/042059

Please Search:

Nucleic Acid and Protein databases for:

SEQ ID NO: 1 and SEQ ID NO: 2

Including:

1. Default Search.
2. Interference Search.

Thanks,
Dave.

3/23/05

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 03:00:36 ; Search time 204 Seconds
(without alignments)
3196.461 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 1504
Sequence: 1 atggacttatgaagccaga...........tcctcctcggctgtactag 843

Scoring table: BLOSUM62
                 Xgapop 10.0 ; Xgapext 0.5
                 Ygapop 10.0 ; Ygapext 0.5
                 Fgapop 6.0 ; Fgapext 7.0
                 Delop   6.0 ; Delext 7.0

Searched: 2105692 seqs, 386760381 residues
                 Listing first 45 summaries

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Total number of hits satisfying chosen parameters: 4211384

Command line parameters:
-MODEL=frame+_n2p:model -DBV=xlh
-Q=/cgn2_1/USPTO_spool/US10042059/runat_30032005_103351_7699/app_query.fasta_1.1031
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10042059 @CGN 1 1 137 @runat_30032005_103351_7699 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:*
    1: geneseqp1908:*
    2: geneseqp1908:*
    3: geneseqp1908:*
    4: geneseqp2001s:*
    5: geneseqp2002s:*
    6: geneseqp2002as:*
    7: geneseqp2003s:*
    8: geneseqp2003bs:*
    8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1434	95.3	280	3	AAB11451	H. polymo
2	921	61.2	268	8	ADP98857	C. albica
3	587.5	39.1	266	8	ADR86097	Aspergill
4	581.5	38.7	263	6	AAO19566	M sterili
5	508.5	33.8	257	6	ABU08093	Rice chor
6	508.5	33.8	257	8	ADJ55264	Soybean c
7	481.5	32.0	316	5	ABB91515	Herbicida
8	481.5	32.0	331	6	ABU08092	Corn chor
9	481.5	32.0	331	8	ADJ55262	Corn chor
10	466	31.0	334	5	ABB92474	Herbicida
11	464	30.9	334	6	ABU08097	A. thalia
12	464	30.9	334	8	ADJ55271	Arabidops
13	423	28.1	261	6	ABU08094	Soybean c
14	423	28.1	261	8	ADJ55266	Soybean c
15	399	26.5	265	5	AAM49654	A. thalia
16	399	26.5	265	8	ABB93424	Herbicida
17	399	26.5	265	6	ABU08098	A. thalia
18	399	26.5	265	8	ADJ55272	Arabidops
19	399	26.5	330	3	AAM49655	Construct
20	391	26.0	319	3	AAG06003	Arabidops
21	391	26.0	347	3	AAG06002	Arabidops
22	391	26.0	207	6	ABU08095	Wheat cho
23	313	20.8	207	8	ADJ55268	Wheat cho
24	313	20.8	154	8	ABU08090	Protein e
25	221	14.7	154	8	ADJ55258	Wheat cho
26	221	14.7	132	8	ABU08089	Protein e
27	197.5	13.1	132	8	ADJ55256	Soybean c
28	197.5	13.1	132	6	ABB98398	Soybean c
29	148	9.8	93	6	ABU08087	Protein e
30	148	9.8	93	8	ADJ55252	Corn chor
31	132	8.8	93	8	ADJ55252	Corn chor
32	132	8.8	248	7	AAR79314	Pseudomon
33	127	8.4	62	8	ABU08088	Protein e
34	122	8.1	62	8	ADJ55254	Rice chor
35	122	8.1	62	8	ABP76678	Streptomy
36	119	7.9	201	4	AAU51115	Propionib
37	119	7.9	201	8	AAU51115	Propionib
38	118	8.0	19938	6	ABP76680	Streptomy
39	118	8.0	19938	6	ABM47634	Propionib
40	132	8.8	19938	6	AAY07202	Mycobacte
41	116	7.7	1093	2	ABP76679	Streptomy
42	115.5	7.7	446	2	AAR41001	Human myo
43	114.5	7.6	469	2	AAY20058	B. burgdo
44	113.5	7.5	539	7	AB078971	Pseudomon
45	113	7.5	409	7	ABO74668	Pseudomon

Abu08097 A. thalia
Adj55271 Arabidops
Abu08094 Soybean c
Adj55266 Soybean c
Aam49654 A. thalia
Abb93424 Herbicida
Abu08098 A. thalia
Adj55272 Arabidops
Aam49655 Construct
Aag06003 Arabidops
Aag06002 Arabidops
Abu08095 Wheat cho
Adj55268 Wheat cho
Abu08090 Protein e
Adj55258 Wheat cho
Abu08089 Protein e
Adj55256 Soybean c
Abb98398 Soybean c
Abu08087 Protein e
Adj55252 Corn chor
Adj55252 Corn chor
Aar79314 Pseudomon
Abu08088 Protein e
Adj55254 Rice chor
Abp76678 Streptomy
Aau51115 Propionib
Aau51115 Propionib
Abp76680 Streptomy
Abm47634 Propionib
Aay07202 Mycobacte
Abp76679 Streptomy
Aar41001 Human myo
Aay20058 B. burgdo
Ab078971 Pseudomon
Abo74668 Pseudomon

ALIGNMENTS

RESULT 1
AAB11451

AAB11451;

ID   AAB11451 standard; protein; 280 AA.
XX
AC   AAB11451;
XX
DT   12-SEP-2003 (revised)
DT   01-MAR-2001 (first entry)
XX
DE   H. polymorpha chorismate mutase protein.
XX
KW   Chorismate mutase; prephenate; selection marker; auxotrophic yeast.
XX
OS   Pichia angusta.
XX
PN   WO200065071-A1.
XX
PD   02-NOV-2000.
XX
PF   27-APR-2000; 2000WO-EP003844.
XX
PR   27-APR-1999; 99DB-01019124.
XX
PA   (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX
PI   Gellissen G, Braus G, Pries R, Krappmann S, Strasser AW;
XX
DR   WPI; 2000-687355/67.
DR   N-PSDB; AAC81949.
XX
PT   Nucleic acids encoding chorismate mutase, useful for preparing an
PT   auxotrophic selection system for recombinant yeast and for recombinant
PT   protein expression.
XX

OM nucleic - protein search, using frame_plus_n2p model

Copyright (c) 1993 - 2005 Compugen ltd.

GenCore version 5.1.6

Run on: March 31, 2005, 04:36:43 ; Search time 99.5 Seconds
(without alignments)
1264.908 Million cell updates/sec

Title:
Perfect score: 1504
Sequence: US-10-042-059B-1
1 atggactttatgaagccaga.........tctcctctggcttgtactag 843

Searched: 513545 seqs, 74649064 residues

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10042059/runat 30032005 103354_7738/app_query.fasta_1.1031
-DB=Issued_Patents_AA -QRMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=gpto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10042059 @CGN 1 1.30 @runat 30032005 103354_7738 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY_NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -LONGLOG
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7 -XGAPEXT=0.5 -FGAPOP=6

Database : Issued_Patents_AA.*
1: /cgn2_6/ptcdata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptcdata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptcdata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptcdata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptcdata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptcdata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	921	61.2	315	4	US-09-248-796A-18213	Sequence 18213, A
2	508.5	33.8	257	4	US-09-454-279-14	Sequence 14, Appl
3	481.5	32.0	331	4	US-09-454-279-12	Sequence 12, Appl
4	464	30.9	334	4	US-09-454-279-21	Sequence 21, Appl
5	423	28.1	261	4	US-09-454-279-16	Sequence 16, Appl
6	399	26.5	265	4	US-09-454-279-22	Sequence 22, Appl
7	313	20.8	207	4	US-09-454-279-18	Sequence 18, Appl
8	221	14.7	154	4	US-09-454-279-8	Sequence 8, Appl
9	197.5	13.1	132	4	US-09-454-279-6	Sequence 6, Appl
10	132	8.8	93	4	US-09-454-279-2	Sequence 2, Appl
11	127	8.4	248	4	US-09-252-991A-28060	Sequence 28060, A
12	122	8.1	62	4	US-09-454-279-4	Sequence 4, Appl
13	113.5	7.5	539	4	US-09-252-991A-27717	Sequence 27717, A
14	113	7.5	409	4	US-09-252-991A-23414	Sequence 23414, A
15	111	7.4	273	4	US-09-252-991A-22218	Sequence 22218, A
16	111	7.4	373	4	US-09-252-991A-28703	Sequence 28703, A
17	109.5	7.3	335	4	US-09-252-991A-23948	Sequence 23948, A
18	109.5	7.3	423	5	PCT-US91-09055-2	Sequence 2, Appli
19	109	7.2	163	4	US-09-252-991A-29867	Sequence 29867, A
20	108	7.2	163	4	US-09-252-991A-24968	Sequence 24968, A
21	108	7.2	2150	5	US-09-321-987B-2	Sequence 2, Appli
22	108	7.2	1480	3	US-09-182-024A-5	Sequence 5, Appli
23	108	7.2	1480	3	US-09-540-153-7	Sequence 7, Appli
24	108	7.2	1480	3	US-09-540-245A-7	Sequence 7, Appli
25	107.5	7.1	296	4	US-09-191-647-7	Sequence 7, Appli
26	107.5	7.1	2150	4	US-09-800-729-155	Sequence 155, App
27	107.5	7.1	2165	4	US-09-321-987B-2	Sequence 2, Appli
28	107	7.1	375	4	US-09-252-991A-24278	Sequence 24278, A
29	107	7.1	697	4	US-09-252-991A-21106	Sequence 21106, A
30	106.5	7.1	589	4	US-09-252-991A-20790	Sequence 20790, A
31	106	7.0	589	4	US-09-252-991A-18210	Sequence 18210, A
32	106	7.0	659	4	US-09-252-991A-25283	Sequence 25283, A
33	105.5	7.0	383	4	US-09-252-991A-29621	Sequence 29621, A
34	105.5	7.0	565	4	US-09-252-991A-17762	Sequence 17762, A
35	105	7.0	583	4	US-09-252-991A-21951	Sequence 21951, A
36	105	7.0	661	4	US-09-252-991A-28056	Sequence 28056, A
37	104.5	7.0	343	4	US-09-252-991A-17876	Sequence 17876, A
38	104.5	6.9	532	4	US-09-252-991A-28042	Sequence 28042, A
39	104	6.9	205	4	US-09-252-991A-28770	Sequence 28770, A
40	104	6.9	208	4	US-09-252-991A-25609	Sequence 25609, A
41	104	6.9	1523	4	US-09-538-032-955	Sequence 955, App
42	103.5	6.9	299	4	US-09-252-991A-21951	Sequence 21951, A
43	103.5	6.9	299	4	US-09-252-991A-17878	Sequence 17878, A
44	103.5	6.9	437	4	US-09-252-991A-32814	Sequence 32814, A
45	103	6.8	266	4	US-09-252-991A-19128	Sequence 19128, A

ALIGNMENTS

RESULT 1
US-09-248-796A-18213
Sequence 18213, Application US/09248796A
Patent No. 6747137.
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 18213
LENGTH: 315
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-18213

Alignment Scores:
Pred. No.: 4.76e-89
Score: 921.00
Percent Similarity: 82.63%
Best Local Similarity: 65.64%
Query Match: 61.24%

Length: 315
Matches: 170
Conservative: 44
Mismatches: 45
Indels: 0
Gaps: 0

US-10-042-059B-1 (1-843) x US-09-248-796A-18213 (1-315)

```
Qy   1 ATGGACTTTATGAAGCCAGAAACAGTGCTTGGACCTTGGCAACATTAGAGATGCCTTGGTC 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  48 MetAspPheMetLysProGluThrValLeuValAlaAsnIleArgGlnAlaLeuVal 67
```

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 05:46:38 ; Search time 275 Seconds
(without alignments)
2032.966 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 1504
Sequence: 1 atggacttatgaagccaga..............tcccctccggcttgtactag 843

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1413372 seqs, 331592847 residues

Total number of hits satisfying chosen parameters: 2826744

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10042059/runat_30032005_103356_7783/app_query.fasta_1.1031
-DB=Published_Applications_AA_.QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10042059 @CGN_1_103_@runat_30032005_103356_7783
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptcdata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptcdata/1/pubpaa/pct_NEW_PUB.pep.*
3: /cgn2_6/ptcdata/1/pubpaa/US09_NEW_PUB.pep.*
4: /cgn2_6/ptcdata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptcdata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptcdata/1/pubpaa/pctUS_PUBCOMB.pep.*
7: /cgn2_6/ptcdata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptcdata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptcdata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptcdata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptcdata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptcdata/1/pubpaa/US10A_PUBCOMB.pep.*
13: /cgn2_6/ptcdata/1/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptcdata/1/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptcdata/1/pubpaa/US10D_PUBCOMB.pep.*
16: /cgn2_6/ptcdata/1/pubpaa/US11_NEW_PUB.pep.*
17: /cgn2_6/ptcdata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptcdata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptcdata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptcdata/1/pubpaa/US60_PUBCOMB.pep.*

Result %   SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	1434	95.3	280	13	US-10-042-059A-2	Sequence 2, Appli
2	921	61.2	268	17	US-10-741-849-7032	Sequence 7032, Ap
3	508.5	33.8	257	9	US-09-454-279-14	Sequence 14, Appl
4	508.5	33.8	334	15	US-10-624-061-14	Sequence 14, Appl
5	508.5	33.8	313	16	US-10-437-963-196439	Sequence 196439,
6	497.5	33.1	320	15	US-10-425-114-68586	Sequence 68586, A
7	495.5	32.9	315	15	US-10-424-599-266662	Sequence 266662,
8	486	32.3	350	16	US-10-767-701-45917	Sequence 45917, A
9	481.5	32.0	331	15	US-10-624-061-12	Sequence 12, Appl
10	481.5	32.0	331	15	US-10-624-061-12	Sequence 12, Appl
11	464	30.9	334	15	US-09-454-279-21	Sequence 21, Appl
12	464	30.9	257	9	US-09-454-279-21	Sequence 21, Appl
13	423	28.1	261	15	US-10-624-061-16	Sequence 16, Appl
14	423	28.1	261	15	US-10-624-061-16	Sequence 16, Appl
15	423	28.1	297	15	US-10-425-114-48703	Sequence 48703, A
16	419	27.9	261	15	US-10-424-599-235878	Sequence 235878,
17	408.5	27.2	255	16	US-10-437-963-151855	Sequence 151855,
18	399	26.5	265	9	US-09-454-279-22	Sequence 22, Appl
19	399	26.5	265	9	US-09-454-279-22	Sequence 22, Appl
20	374.5	24.9	313	15	US-10-624-061-22	Sequence 22, Appl
21	373	24.8	264	15	US-10-425-114-65608	Sequence 65608, A
22	356	23.7	306	15	US-10-425-114-66528	Sequence 66528, A
23	348	23.1	316	16	US-10-437-963-186047	Sequence 186047,
24	344.5	22.9	186	15	US-10-437-963-186040	Sequence 186040,
25	328	21.8	199	15	US-10-425-114-52677	Sequence 52677, A
26	313	20.8	207	9	US-09-454-279-18	Sequence 18, Appl
27	313	20.8	207	15	US-10-425-114-43606	Sequence 43606, A
28	297.5	19.8	489	16	US-10-437-963-103185	Sequence 103185,
29	282	18.8	210	16	US-10-767-701-32934	Sequence 32934, A
30	221	14.7	154	9	US-09-454-061-8	Sequence 8, Appli
31	221	14.7	154	9	US-09-454-061-8	Sequence 8, Appli
32	197.5	13.1	132	15	US-10-624-061-6	Sequence 6, Appli
33	197.5	13.1	132	15	US-10-624-061-6	Sequence 6, Appli
34	178	11.8	109	15	US-10-424-599-224668	Sequence 224668,
35	157.5	10.5	97	15	US-10-425-114-38176	Sequence 38176, A
36	151.5	10.1	19695	9	US-10-084-846A-3	Sequence 3, Appli
37	132	8.8	93	15	US-10-624-061-2	Sequence 2, Appli
38	132	8.8	93	15	US-10-624-061-2	Sequence 2, Appli
39	130.5	8.9	278	15	US-10-425-114-44480	Sequence 44480, A
40	130.5	8.9	279	15	US-10-425-114-43408	Sequence 43408, A
41	124.5	8.3	797	14	US-10-156-761-10907	Sequence 10907, A
42	122	8.1	62	15	US-09-454-279-4	Sequence 4, Appli
43	122	8.1	62	15	US-10-437-963-116837	Sequence 116837,
44	119	7.9	182	16	US-10-624-061-116837	Sequence 116837,
45	117.5	7.8	459	16	US-10-437-963-153917	Sequence 153917,

Result   Query   ALIGNMENTS

RESULT 1
US-10-042-059A-2
; Sequence 2, Application US/10042059A
; Publication No. US20020197704A1
; GENERAL INFORMATION:
; APPLICANT: Rhein Biotech Gesellschaft fur neue biotechnologische Prozesse und
; APPLICANT: Produkte mbH
; TITLE OF INVENTION: Nucleic Acid Molecule Containing a Nucleic Acid Coding for a Pol
; TITLE OF INVENTION: with Chorismate Mutase Activity
; FILE REFERENCE: PA20558US-019
; CURRENT APPLICATION NUMBER: US/10/042,059A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: DE 199 19 124.7
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Hansenula polymorpha
US-10-042-059A-2

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run On: March 31, 2005, 04:14:20 ; Search time 44 Seconds
(without alignments)
3686.851 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 1504
Sequence: 1 atggacttatgaagccaga..........tctcctggcttgtactag 843

Scoring table:
BLOSUM62
Xgapop 10.0   Xgapext 0.5
Ygapop 10.0   Ygapext 0.5
Fgapop 6.0   Fgapext 7.0
Delop 6.0   Delext 7.0

Searched: 283416 seqs, 96216763 residues
Listing first 45 summaries

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Command line parameters:
-MODEL=frame+_n2p:model -DEV=xlh
-Q=/ccpn2.1/USPTO_spool/US10042059/runat_30032005_103353_7727/app_query.fasta_1.1031
-DB=PIR_79 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=20000000000
-USER=US10042059 @CGN_1_1.25 @runat_30032005_103353_7727-NCPU=6 -ICPU=3
-NO_MMAP=-LARGEQUERY_NEG_SCORES0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	776.5	51.6	256	2	A45921	chorismate mutase
2	558.5	37.1	251	2	T37715	probable chorismat
3	464	30.9	334	2	T32739	chorismate mutase
4	399	26.5	265	2	S39958	chorismate mutase
5	166	11.0	82	2	T50796	chorismate mutase
6	119	7.9	70	2	T14902	chorismate mutase
7	114.5	7.6	469	2	T14901	hypothetical prote
8	108	7.1	1480	2	C70109	slit protein 1 pre
9	107.5	7.1	2165	2	A36665	hypothetical prote
10	104	6.9	313	2	T21171	tetrahydropteroylt
11	104	6.9	336	2	B75138	site-specific DNA-
12	104	6.9	1203	2	S04739	DNA-binding protei
13	103.5	6.9	1732	2	S26650	protein kinase (EC
14	103	6.8	560	1	T14039	arginine-tRNA liga
					F69059	
c	15	102.5	6.8	376	2	P75503	conserved hypothet
16	102	6.8	1033	2	T37711	actin-interacting
17	101.5	6.7	625	2	T37715	hypothetical prote
18	101.5	6.7	728	2	T32739	hypothetical prote
19	99.5	6.6	534	2	S26427	70K structural pro
20	99	6.6	2142	1	T39903	serine-rich protei
21	98.5	6.5	424	2	ZLVNPV	genome polyprotein
22	98	6.5	1702	2	T14050	hypothetical prote
23	98	6.5	1791	2	H87520	protein kinase (EC
24	97.5	6.5	415	2	T02345	hypothetical prote
25	97.5	6.5	492	2	D95275	probable two-compo
26	97.5	6.5	1286	2	A87471	hypothetical proce
27	97	6.4	323	2	T37366	RNA polymerase sub
28	97	6.4	514	1	D90527	glycerol-3-phospha
29	97	6.4	1167	2	W2WL5	E2 protein - human
c	30	96.5	6.4	829	2	E96963	DNA polymerase III
31	96.5	6.4	932	2	S72366	DNA topoisomerase
32	96.5	6.4	2232	2	F84465	hypothetical proce
33	96	6.4	528	2	T34434	hypothetical prote
34	96	6.4	680	2	B75310	conserved hypothet
35	96	6.4	808	2	S21323	probable endogluca
c	36	96	6.4	1286	2	H64474	hypothetical prote
37	96	6.4	1286	2	H36845	DNA-directed RNA p
38	96	6.4	1286	2	T28521	DNA-directed RNA p
39	96	6.4	2127	2	A72161	M6R protein - vari
40	95.5	6.3	280	2	H72593	genome polyprotein
41	95	6.3	623	2	S55652	hypothetical prote
42	94.5	6.3	310	2	S43865	transcription cont
c	43	94.5	6.3	402	2	E86185	cytokeratin 8, typ
44	94.5	6.3	790	2	T19683	hypothetical prote
45	94.5	6.3	1287	1	RNVZ47	DNA-directed RNA p

ALIGNMENTS

RESULT 1
A45921
chorismate mutase (EC 5.4.99.5) - Yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YP9499.15c; protein YPR060c
C;Species: Saccharomyces cerevisiae
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C;Accession: A45921; S54081
R;Schmidheini, T.; Sperisen, P.; Paravicini, G.; Huetter, R.; Braus, G.
J. Bacteriol. 171; 1245-1253, 1989
A;Title: A single point mutation results in a constitutively activated and feedback-res
A;Reference number: A45921; MUID:89155418; PMID:2646272
A;Accession: A45921
A;Molecule type: DNA
A;Residues: 1-256 <SCH>
A;Cross-references: UNIPROT:P32178; GB:M24517; NID:g295576; PIDN:AAB59309.1; PID:g2955
R;Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54059
A;Accession: S54081
A;Molecule type: DNA
A;Residues: 1-256 <BAD>
A;Cross-references: EMBL:Z49219; NID:g805025; PIDN:CAA89177.1; PID:g805040; MIPS:YPR06C
C;Genetics:
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:ARO7
A;Cross-references: SGD:S0006264; MIPS:YPR060c
A;Map position: 16R
C;Function:
A;Description: intramolecular transferase; isomerase
A;Pathway: aromatic amino acid biosynthesis
C;Superfamily: chorismate mutase of the AroQ class, eukaryotic type
C;Keywords: aromatic amino acid biosynthesis; intramolecular transferase; isomerase

Alignment Scores:
Pred. No.: 6.12e-55
Score: 776.50
Percent Similarity: 74.13%
Best Local Similarity: 58.69%

Length: 256
Matches: 152
Conservative: 40
Mismatches: 62

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 31, 2005, 03:04:01 ; Search time 247.5 Seconds
(without alignments)
3488.345 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 1504
Sequence: 1 atggacttatgaagccaga..........tctcctctggcttgtactag 843

Scoring table: BLOSUM62
                Xgapop 10.0 ; Xgapext 0.5
                Ygapop 10.0 ; Ygapext 0.5
                Fgapop 6.0 ; Fgapext 7.0
                Delop   6.0 ; Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p:model1 -DEV=xlh
-Q=/cgn2_1/USPTO_spool/USI10042059/runat_30032005_103352_7710/app_query.fasta_1.1031
-DB=UniProt_03 -QFMT=faatan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MIN=0 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=USI10042059 @CGN 1 1 140 @runat_30032005_103352_7710 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_03:*
            1: uniprot_sprot:*
            2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1434	95.3	280	2	Q9P4D8	pichia angu
2	907	60.3	267	2	Q6BK64	debaryomyce
3	808	53.7	267	2	Q6CS37	yarrowia li
4	776.5	51.6	256	1	CHMU_YEAST	saccharomyc
5	752.5	50.0	260	2	Q6CVY3	kluyveromyc
6	719	47.8	260	2	Q6FLZ7	candida gla
7	702.5	46.7	259	2	Q75BG5	ashbya goss
8	594	39.5	259	2	Q7SBR4	neurospora
9	581.5	38.7	263	2	Q6L8Q0	roseellinia
10	579	38.5	267	2	Q9Y7B2	emericella
11	558.5	37.1	251	1	CHMU_SCHPO	schizosacch
12	508.5	33.8	294	2	Q93Y60	oryza sativ
13	481.5	32.0	316	2	Q9C544	arabidopsis
14	475	31.6	295	2	Q96VZ8	cryptococcu
15	466	31.0	334	1	CHMU_ARATH	arabidopsis
16	465	30.9	255	2	Q9STB2	lycopersico
17	464.5	30.9	316	2	Q9XF60	arabidopsis
18	408.5	27.2	255	2	Q6H819	oryza sativ
19	399	26.5	265	2	Q9S7H4	arabidopsis
20	391	26.0	284	2	Q629E6	oryza sativ
21	290.5	19.3	179	2	Q6F3R1	oryza sativ
22	247	16.4	131	2	Q65861	nicotiana t
23	166	11.0	82	2	Q65861	prunus aviu
24	119	7.9	70	2	O22410	petroselinu
25	114.5	7.7	1225	2	O22409	petroselinu
26	111	7.4	637	2	O51102	petroselinu
27	111	7.4	822	2	Q6DFC7	borrelia bu
28	110.5	7.3	2607	2	Q8BT18	bacillus li
29	110.5	7.3	492	2	O84JR3	xenopus lae
30	109.5	7.3	683	2	Q84JR3	mus musculu
31	109	7.4	1236	2	O9C105	arabidopsis
32	108	7.2	1504	1	Q9C105	schizosacch
33	107.5	7.1	2165	2	P24014	drosophila
34	107	7.1	622	2	Q19791	caenorhabdi
35	106.5	7.1	413	2	Q7NAK6	mycoplasma
36	106.5	7.1	683	2	Q8R3P7	mus musculu
37	106.5	7.1	1158	2	Q9DWA5	rat cytomeg
38	106.5	7.1	1286	2	Q9CSS1	mus musculu
39	106.5	7.1	1371	2	Q8V4Z4	monkeypox v
40	106.5	7.1	1461	2	Q6PDI7	mus musculu
41	106.5	7.1	2404	1	Q6ZPE9	mus musculu
42	106	7.0	5289	2	Q9QX47	mus musculu
43	105.5	7.0	317	2	O88Z03	lactobacill
44	105.5	7.0	1638	2	Q99697	homo sapien
45	105	7.0	600	2	Q8IWQ7	homo sapien

ALIGNMENTS

RESULT 1
Q9P4D8

ID   Q9P4D8   PRELIMINARY;   PRT;   280 AA.
AC   Q9P4D8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Chorismate mutase (EC 5.4.99.5).
GN   Name=HARO7;
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4905;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB11;
RX   MEDLINE=20353451; PubMed=10994726;
RX   DOI=10.1128/JB.182.15.4188-4197.2000;
RA   Krappmann S., Pries R., Gellissen G., Hiller M., Braus G.H.;
RT   "HARO7 encodes chorismate mutase of the methylotrophic yeast Hansenula
RT   polymorpha and is derepressed upon methanol utilization.";
RL   J. Bacteriol. 182:4188-4197(2000).
DR   EMBL; AF204738; AAF87954.1; -.
DR   HSSP; P32178; 2CSM.
DR   GO; GO:0004106; F:chorismate mutase activity; IEA.
DR   GO; GO:0016853; F:isomerase activity; IEA.
DR   GO; GO:0009073; P:aromatic amino acid family biosynthesis; IEA.
DR   InterPro; IPR002701; Chorismate_mut.
DR   InterPro; IPR008238; Chor_mut_AroQ_eu.
DR   Pfam; PF01817; CM_2; 1.
DR   PIRSF; PIRSF017318; Chor_mut_AroQ_eu; 1.
DR   TIGRFAMs; TIGR01802; CM_pl-yst; 1.
KW   Isomerase.
SQ   SEQUENCE   280 AA;   32068 MW;   689D49312CD292D9 CRC64;

Alignment Scores:
Pred. No.:   6.99e-108
Score: 1434.00   Length: 280
Percent Similarity: 100.00%   Matches: 280
Best Local Similarity: 100.00%   Conservative: 0
                                        Mismatches: 0

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: March 30, 2005, 21:05:55
Search time 4115 Seconds
(without alignments)
9926.548 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 1 atggacttatgaagccaga.........tctcctctggccttgtactag 843

Scoring table: IDENTITY NUC
                Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422707955 residues

Total number of hits satisfying chosen parameters: 9416466

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Database : GenEmbl:*
   1: gb_ba:*
   2: gb_htg:*
   3: gb_in:*
   4: gb_om:*
   5: gb_ov:*
   6: gb_pat:*
   7: gb_ph:*
   8: gb_pl:*
   9: gb_pr:*
  10: gb_ro:*
  11: gb_sts:*
  12: gb_sy:*
  13: gb_un:*
  14: gb_vi:*

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	DB ID	Description
	1	843	100.0	843	6	BD266599	BD266599 Nucleic a
	2	843	100.0	1652	8	AX043758	AX043758 Sequence
	3	843	100.0	1652	8	AF204738	AF204738 Pichia an
	4	843	100.0	1655	6	BD266600	BD266600 Nucleic a
	5	843	100.0	1655	6	AX043760	AX043760 Sequence
c	6	325.2	38.6	110000	8	CR382131_20	CR382131_20 Continuation (21 o
c	7	281.8	33.4	110000	8	CR382138_20	CR382138_20 Continuation (21 o
	8	270	32.0	948	8	AR548979	AR548979 Sequence
c	9	256.4	30.4	110000	8	AB016817_01	AB016817_01 Continuation (2 of
c	10	217.4	25.8	2059	8	YSCARO7A	YSCARO7A Saccharomyc
c	11	217.4	25.8	2059	8	SC9499X	SC9499X S.cerevisia
c	12	217.4	25.8	165536	8	SCCHRXVI	SCCHRXVI Saccharomyc
c	13	217.2	25.8	771	8	AY693179	AY693179 Saccharom
c	14	210.4	25.0	110000	8	CR382122_07	CR382122_07 Continuation (8 of
c	15	189	22.4	110000	8	CR380957_11	CR380957_11 Continuation (12 o
	16	156.4	18.6	869	6	BD178335	BD178335 Transform
	17	156.4	18.6	869	6	AB116236	AB116236 Rosellini
	18	141.8	16.8	1348	8	AK105512	AK105512 Oryza sat
	19	131.4	15.6	1231	6	AR404687	AR404687 Sequence
	20	131.4	15.6	1250	8	AK068983	AK068983 Oryza sat
	21	113.6	13.5	3167	8	AF133241	AF133241 Emericell
	22	111.2	13.2	1223	8	AR404686	AR404686 Sequence
	23	110	13.0	1000	8	L47356	L47356 Lycopersico
	24	107.6	12.8	951	8	BT005306	BT005306 Arabidops
	25	107.6	12.8	1143	8	AK117860	AK117860 Arabidops
	26	104.4	12.4	1217	6	AR236635	AR236635 Sequence
c	27	104.4	12.4	1217	6	AR236641	AR236641 Sequence
c	28	104.4	12.4	1217	6	AX464576	AX464576 Sequence
c	29	104.4	12.4	1217	6	AX464582	AX464582 Sequence
	30	100.6	11.9	110000	2	AP006499_05	AP006499_05 Continuation (6 of
	31	100.0	11.8	1020	6	AF131219	AF131219 Arabidops
	32	99.8	11.8	798	6	AX506277	AX506277 Sequence
	33	97.4	11.6	829	6	AY133840	AY133840 Arabidops
	34	97.4	11.6	993	6	AX343935	AX343935 Sequence
	35	97.4	11.6	1006	6	AX343933	AX343933 Sequence
	36	97.4	11.6	1006	6	AR236634	AR236634 Sequence
	37	97.4	11.6	1006	6	AR236640	AR236640 Sequence
	38	97.4	11.6	1006	6	AY133840	AY133840 Arabidops
	39	97.4	11.6	1006	6	AX343935	AX343935 Sequence
c	40	97.4	11.6	1006	6	AX464575	AX464575 Sequence
c	41	97.4	11.6	1006	8	AX464581	AX464581 Sequence
c	42	97.4	11.6	1055	8	L47355	L47355 Arabidops
	43	93.6	11.1	1005	8	AY065238	AY065238 Arabidops
	44	93.6	11.1	1207	6	AX506042	AX506042 Sequence
c	45	93.6	11.1	1207	6	AR236633	AR236633 Sequence
						AR236639	AR236639 Sequence

ALIGNMENTS

RESULT 1
BD266599
BD266599

```
LOCUS       BD266599                 843 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Nucleic acid molecule, containing a nucleic acid which codes for a
            polypeptide with chorismate mutase activity.
ACCESSION   BD266599
VERSION     BD266599.1  GI:33076367
KEYWORDS    JP 2002542788-A/1.
SOURCE      Pichia angusta
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Gellissen,G., Braus,G., Pries,R., Krappmann,S. and Strasser,A.W.
  TITLE     Nucleic acid molecule, containing a nucleic acid which codes for a
            polypeptide with chorismate mutase activity
  JOURNAL   Patent: JP 2002542788-A 1 17-DEC-2002;
            RHEIN BIOTECH GESELLSCHAFT FUER NEUE BIOTECHNOLOGISCHE PROZESSE UND
            PRODUKTE MBH
COMMENT     OS   Hansenula polymorpha (yeast)
            PN   JP 2002542788-A/1
            PD   17-DEC-2002
            PF   27-APR-2000 JP 2000614405
            PR   27-APR-2000 JP 199 19 124.7
            PI   GERD GELLISSEN,GERHARD BRAUS,RALPH PRIES,SVEN KRAPPMANN, PI
            ALEXANDER W STRASSER
            PC   C12N15/09,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
            ,C12N9/90,C12N15/00,
            PC   C12N5/00
            CC   Nucleic acid molecule, containing a nucleic acid which codes
            CC   for a
            CC   polypeptide with chorismate mutase activity
FEATURES             Location/Qualifiers
     source          1..843
     FH   Key             Location/Qualifiers
     FT   source          1..843
     FT                   /organism="Hansenula polymorpha (yeast)"
     FT                   /organism="Pichia angusta"
     FT                   /mol_type="genomic DNA"
     FT                   /db_xref="taxon:4905"
ORIGIN
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 30, 2005, 18:17:40 ; Search time 588 Seconds
                                          (without alignments)
                                          8486.976 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 1 atggaacttatgaagccaga..........tctcctctggctctgctactag 843

Scoring table: IDENTITY_NUC
                Gapop 10.0 ; Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   N_Geneseq_16Dec04:*
                1: geneseqn1908:*
                2: geneseqn1990s:*
                3: geneseqn2000s:*
                4: geneseqn2001as:*
                5: geneseqn2001bs:*
                6: geneseqn2002as:*
                7: geneseqn2002bs:*
                8: geneseqn2003as:*
                9: geneseqn2003bs:*
               10: geneseqn2003cs:*
               11: geneseqn2003ds:*
               12: geneseqn2004as:*
               13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	843	100.0	843	3	AAC81949	Aac81949 H. polymo
2	843	100.0	1655	3	AAC81950	Aac81950 H. polymo
3	270	32.0	807	12	ADP98547	Adp98547 C. albica
4	182.8	21.7	801	13	ADR85510	Adr85510 Aspergill
5	156.4	18.6	869	8	AAL50202	Aal50202 M sterili
6	141.2	16.7	989	13	ADR84923	Adr84923 Aspergill
7	141.2	16.7	2920	8	ABT17891	Abt17891 Aspergill
8	141.2	16.7	2987	13	ABT19705	Abt19705 Aspergill
9	141.2	16.7	6989	13	ADR84336	Adr84336 Aspergill
10	131.4	15.6	1231	10	ABX93057	Abx93057 Rice chor
11	131.4	15.6	1231	12	ADJ55263	Adj55263 Rice chor
12	116.8	13.9	658	10	AAF14469	Aaf14469 Aspergill
13	111.2	13.2	1223	10	ABX93056	Abx93056 Corn chor
14	111.2	13.2	1223	12	ADJ55261	Adj55261 Corn chor
15	111.2	13.2	1217	6	ABA91395	Aba91395 Arabidops
16	104.4	12.4	1217	6	ABA91401	Aba91401 Arabidops
17	99.8	11.8	1020	6	ABX93058	Abx93058 Soybean c
18	99.8	11.8	1020	12	ADJ55265	Adj55265 Soybean c
19	97.4	11.6	798	6	AB213167	Ab213167 Arabidops
20	97.4	11.6	993	6	ABA99614	Aba99614 Construct
21	97.4	11.6	1006	6	ABA99613	Aba99613 A. thalia
22	97.4	11.6	1006	6	ABA91400	Aba91400 Arabidops
23	97.4	11.6	1006	6	ABA91394	Aba91394 Arabidops
24	91.6	11.1	1005	6	AB212932	Ab212932 Arabidops
25	93.6	11.1	1207	6	ABA91393	Aba91393 Arabidops
26	93.6	11.1	1207	6	ABA91399	Aba91399 Arabidops
27	93.6	8.7	1274	3	AAC33324	Aac33324 Arabidops
28	73	8.7	580	13	ACN57510	Acn57510 Cotton gy
29	69.8	8.3	622	13	ACN50132	Acn50132 Cotton no
30	64.2	7.6	600	13	ADR64815	Adr64815 Cotton CD
31	61.8	7.3	258	6	ABL71526	Abl71526 Corn tass
32	61	7.2	780	10	ABX93059	Abx93059 Wheat cho
33	61	7.2	780	12	ADJ55267	Adj55267 Wheat cho
34	59.8	7.1	579	12	ABX93054	Abx93054 Wheat cho
35	59.8	7.1	579	12	ADJ55257	Adj55257 Wheat cho
36	56.2	6.7	584	13	ACN50205	Acn50205 Cotton no
37	54.8	6.5	542	13	ACN57601	Acn57601 Cotton gy
38	53.4	6.3	525	10	ABX93053	Abx93053 Soybean c
39	53.4	6.3	525	12	ADJ55255	Adj55255 Soybean c
40	52.6	6.2	524	13	ADR64235	Adr64235 Cotton CD
41	49.4	5.9	2000	8	ADA71938	Ada71938 Rice gene
42	49.4	5.8	864	12	ADJ44958	Adj44958 Plant cDN
43	47.6	5.6	1379	13	ABA91794	Aba91794 Yeast mit
44	47	5.6	5858	8	AADS5726	Aads5726 Nephila m
45	46.8	5.6	180385	10	ADL13931	Adl13931 Osteoarth

ALIGNMENTS

RESULT 1
AAC81949
ID   AAC81949 standard; DNA; 843 BP.
XX
AC   AAC81949;
XX
DT   15-SEP-2003 (revised)
DT   01-MAR-2001 (first entry)
XX
DE   H. polymorpha chorismate mutase DNA.
XX
KW   Chorismate mutase; prephenate; selection marker; auxotrophic yeast; ds.
XX
OS   Pichia angusta.
XX
PN   WO200065071-A1.
XX
PD   02-NOV-2000.
XX
PF   27-APR-2000; 2000WO-EP003844.
XX
PR   27-APR-1999;   99DE-01019124.
XX
PA   (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX
PI   Gellissen G, Braus G, Pries R, Krappmann S, Strasser AW;
XX
DR   WPI; 2000-687355/67.
DR   P-PSDB; AAB11451.
XX
PT   Nucleic acids encoding chorismate mutase, useful for preparing an
PT   auxotrophic selection system for recombinant yeast and for recombinant
PT   protein expression.
XX
PS   Claim 1a; Page 57; 63pp; German.
XX
CC   This invention describes novel nucleic acids (I) that encode a
CC   polypeptide (II) with chorismate mutase (CM) activity (or its
CC   complementary strand). CM catalyzes conversion of chorismate to
CC   prephenate, an essential precursor for Phe and Tyr. (I) is a selection
CC   marker for construction of corresponding auxotrophic yeast (requiring Phe
CC   and Tyr) that are used for recombinant production of proteins. (I) allows
CC   selection of transformed yeast on simple media. (Updated on 15-SEP-2003

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 00:11:41 ; Search time 195 Seconds
                (without alignments)
                7073.750 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 1 atggacttatgaagccaga.........tctcctctggcttgactag 843
Scoring table: IDENTITY NUC
                Gapop 10.0 , Gapext 1.0
Searched: 1202784 seqs, 81813B359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%   Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents NA:*
                1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5: /cgn2_6/ptodata/1/ina/pCTUS_COMB.seq:*
                6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

                Pred. No. is the number of results predicted by chance to have a
                score greater than or equal to the score of the result being printed,
                and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	270	32.0	948	4	US-09-248-796A-4110	Sequence 4110, Ap
	2	131.4	15.6	1231	4	US-09-454-279-13	Sequence 13, Appl
	3	111.2	13.2	1223	4	US-09-454-279-11	Sequence 11, Appl
	4	104.4	12.4	1217	3	US-09-610-040-4	Sequence 4, Appl
c	5	104.4	12.4	1217	4	US-09-610-040-10	Sequence 10, Appl
c	6	104.4	12.4	1217	3	US-10-267-763-4	Sequence 4, Appl
c	7	104.4	12.4	1217	4	US-10-267-763-10	Sequence 10, Appl
c	8	99.8	11.8	1020	4	US-09-454-279-15	Sequence 15, Appl
c	9	97.4	11.6	1006	3	US-09-610-040-9	Sequence 9, Appli
c	10	97.4	11.6	1006	4	US-09-610-040-3	Sequence 3, Appli
c	11	97.4	11.6	1006	4	US-10-267-763-3	Sequence 3, Appli
c	12	97.4	11.6	1006	3	US-10-267-763-9	Sequence 9, Appli
c	13	93.6	11.1	1207	4	US-09-610-040-2	Sequence 2, Appli
c	14	93.6	11.1	1207	3	US-09-610-040-8	Sequence 8, Appli
c	15	93.6	11.1	1207	4	US-10-267-763-2	Sequence 2, Appli
c	16	93.6	11.1	1207	3	US-10-267-763-8	Sequence 8, Appli
c	17	61	7.2	780	4	US-09-454-279-17	Sequence 17, Appl
c	18	60.8	7.2	718	1	US-08-232-463-14	Sequence 14, Appl
c	19	59.8	7.1	579	4	US-09-454-279-7	Sequence 7, Appl
c	20	53.4	6.3	525	4	US-09-454-279-5	Sequence 5, Appl
c	21	37.6	4.5	4079	4	US-09-016-434-1248	Sequence 1248, Ap
c	22	37.6	4.5	4517	4	US-09-949-016-4573	Sequence 4573, Ap
c	23	37.6	4.5	4519	4	US-09-023-655-1202	Sequence 1202, Ap
c	24	37.6	4.5	4519	4	US-09-492-027-3	Sequence 650, App
c	25	37.6	4.5	4519	4	US-09-252-991A-15307	Sequence 15307, A
c	26	37.2	4.4	1590	4	US-09-252-991A-15532	Sequence 15532, A
c	27	37.2	4.4	2352	4	US-09-252-991A-15532	Sequence 15532, A
	28	37.2	4.4	2916	4	US-09-252-991A-15259	Sequence 15259, A
	29	36.8	4.4	2169	3	US-09-103-840A-3	Sequence 3, Appli
	30	36.6	4.3	5176	3	US-09-434-408-3	Sequence 3, Appli
	31	36.6	4.3	5176	4	US-09-610-040-6	Sequence 6, Appli
	32	36.4	4.3	601	4	US-10-267-763-6	Sequence 6, Appli
c	33	35.4	4.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	34	35.4	4.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	35	35	4.2	2097	4	US-09-252-991A-13635	Sequence 13635, A
c	36	35	4.2	2331	4	US-09-252-991A-13809	Sequence 13809, A
c	37	34.6	4.1	789	4	US-09-252-991A-12701	Sequence 12701, A
c	38	34.6	4.1	1632	4	US-09-252-991A-13259	Sequence 13259, A
c	39	34.4	4.1	3008	3	US-09-435-376-6	Sequence 6, Appli
c	40	34.2	4.1	1509	4	US-09-724-797-89	Sequence 89, Appli
c	41	34.2	4.1	11958	3	US-09-134-246-8	Sequence 8, Appli
c	42	34.2	4.1	11958	4	US-09-664-186-8	Sequence 8, Appli
c	43	34	4.0	705	4	US-09-270-767-5061	Sequence 5061, Ap
c	44	34	4.0	705	4	US-09-270-767-20343	Sequence 20343, A
c	45	34	4.0	978	4	US-09-902-540-6442	Sequence 6442, Ap

ALIGNMENTS

RESULT 1
US-09-248-796A-4110
; Sequence 4110, Application US/0924B796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4110
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4110

Query Match   32.0%; Score 270; DB 4; Length 948;
Best Local Similarity   59.1%;
Matches 462; Conservative   0; Mismatches 320; Indels   0; Gaps   0;

```
Qy      1 ATGGACTTTATGAAGCCAGAAAACAGTGCTGGACCTTGGCAAACATTAGAGATGCCTTGGTC 60
          |||||| ||||| |||||||||| | |||||||| |||| ||||||||  | ||||| |||||
Db    142 ATGGATTTTATGAAAACCAGAAAACTGTCGTTGCTTGGACTTGGTCAGACATTGGTA 201

Qy     61 CGGGATGGAGGAGATACGAGATCATTCTTCAACTTTATTCGAGCGCGTCGCAGTTCTGTCGCCC 120
          ||| ||||||||| ||||  |||| | ||| | ||| |||  | ||| ||| |||| ||| |||
Db    202 AGGATGGAGGAAGATACTATTCTGTTTGTATTTTAATCGAAAAGATCTCAAATTTTTCAGTCCCCA 261

Qy    121 TCGGTATATACAAAAGTCAACCAGTTCCCTATTCCCAACTTCGACGGCTTCGTTCTTGGACTGG 180
          || ||| ||| |||||  |||   |||  |||| | ||| ||||| ||| |||  |||| |||
Db    262 TCAGTTATTTATGAAAAGAAATAAATATATAATACCCCCAATTTTTGATGGAAACTTTTTTTGGAATGG 321

Qy    181 CTGTTGTCGCAGCACGCAGCGGGAAATCCATTCGTTCCAGTGTGAGGAGATACGACGCGCCACGAGCGAG 240
          ||| ||| ||| | || || |||||||| |||| | |  |   ||||| ||| | ||| |||| ||| ||
Db    322 GCTTTGTTACAAATGGAGAGTTGCTCATTCTCAAATCAGACGGTTATGAAGGCACCAGACGAA 381

Qy    241 GTGGCCTTTTTTTTCCCCAACGTGCTGGAAAAAAACGTTTCTTCGCCCAAGATCAACTACCCATCG 300
          ||| || ||||||| | || ||  | |||||||||  |||  ||||| |||||||||| |||| ||
Db    382 ACTCCATTTTTTCCAGATCAAATGAAAAACTCCAATTTTTACCCGCACAATCAATTACCCTAA 441

Qy    301 GTGCTAGCCTCCTACGCGGCGGATGAAATCAACGTCAACAAAGAGATTACTCAAGATCTTACACG 360
          || |||||| |||  | | |||| |||||| ||  |||| |||||||| | ||| |||| | | |
Db    442 ATATTGGGCCAAAATATTCTGATGAAATTAAGTTAATTCTGAAAATAATGATGAAGTTTTATGTT 501
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 02:00:01 ; Search time 3261 Seconds
                                             (without alignments)
                                             1564.628 Million cell updates/sec

Title:
Perfect score: 843
Sequence: 1 atggacttatgaagccaga...........tctcctcctggcttgtactag 843

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 5607317 seqs, 302245999 residues

Total number of hits satisfying chosen parameters: 11214634

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications NA:*
                 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 3: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                 7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                 17: /cgn2_6/ptodata/1/pubpna/US10D_NEW_PUB.seq:*
                 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
                 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Scoring: US-10-042-059B-1

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	843	100.0	843	13	US-10-042-059A-1	Sequence 1, Appli
2	843	100.0	1655	13	US-10-042-059A-3	Sequence 3, Appli
3	270	32.0	807	19	US-10-741-849-6032	Sequence 6032, Ap
4	156.4	18.6	869	18	US-10-472-587-26	Sequence 26, Appli
c	5	141.2	16.7	2920	15	US-10-128-714-249	Sequence 249, App
c	6	141.2	16.7	2987	15	US-10-128-714-5249	Sequence 5249, Ap
7	131.4	15.6	1016	18	US-10-437-963-93956	Sequence 93956, A
8	131.4	15.6	1231	9	Sequence 13, Appl		
9	131.4	15.6	1231	17	US-10-624-061-13	Sequence 13, Appl
10	118.2	14.0	1423	17	US-10-424-599-123820	Sequence 123820,
11	116.8	13.9	658	18	US-10-653-047-6992	Sequence 6992, Ap
12	116.6	13.8	1146	17	US-10-425-114-7285	Sequence 7285, Ap
13	116	13.8	1376	18	US-10-425-115-102347	Sequence 102347,
14	112.2	13.3	1548	18	US-10-739-930-4455	Sequence 4455, Ap
15	111.2	13.2	1223	9	US-09-454-279-11	Sequence 11, Appl
16	111.2	13.2	1223	17	US-10-624-061-11	Sequence 11, Appl
17	111.2	13.2	1500	18	US-10-767-701-14353	Sequence 14353, A
18	110.6	13.1	1010	18	US-10-425-115-102351	Sequence 102351,
19	109.6	13.0	1397	18	US-10-425-115-102346	Sequence 102346,
20	104.4	12.4	1217	18	US-10-267-763-10	Sequence 10, Appli
21	104.4	12.4	1217	14	US-10-267-763-4	Sequence 4, Appli
22	103.2	12.2	1368	23	US-10-425-115-102349	Sequence 102349,
23	99.8	11.8	1020	14	US-09-938-842A-737	Sequence 737, App
24	99.8	11.8	1020	18	US-10-425-114-7639	Sequence 7639, Ap
25	99.8	11.8	1079	17	US-10-425-114-7639	Sequence 7639, Ap
26	97.4	11.6	798	11	US-09-454-279-11	Sequence 11, Appli
27	97.4	11.6	798	17	US-10-624-061-11	Sequence 11, Appli
28	97.4	11.6	1006	14	US-10-267-763-3	Sequence 3, Appli
c	29	97.4	11.6	1006	14	US-10-267-763-9	Sequence 9, Appli
30	97	11.5	1430	18	US-10-739-930-4454	Sequence 4454, Ap
31	93.6	11.1	1005	9	US-09-938-842A-737	Sequence 737, App
32	93.6	11.1	1005	11	US-09-938-842A-737	Sequence 737, App
33	93.6	11.1	1207	14	US-10-267-763-2	Sequence 2, Appli
c	34	93.6	11.1	1207	18	US-10-267-763-8	Sequence 8, Appli
c	35	93.6	11.1	1299	18	US-10-739-930-621	Sequence 621, App
36	87.6	10.4	1653	17	US-10-424-599-93036	Sequence 93036, A
37	81	9.6	880	14	US-10-425-114-3898	Sequence 3898, Ap
38	80	9.5	1314	18	US-10-425-115-151470	Sequence 151470,
39	78.4	9.3	1197	17	US-10-425-114-27576	Sequence 27576, A
40	77.2	9.2	524	11	US-10-425-115-102342	Sequence 102342,
41	76.2	9.0	768	14	US-10-437-963-49372	Sequence 49372, A
42	74.8	8.9	2154	18	US-10-437-963-83557	Sequence 83557, A
43	74.2	8.8	1066	17	US-10-425-114-1332	Sequence 1332, Ap
44	74.2	8.8	1083	18	US-10-425-115-1155559	Sequence 1155559,
45	73	8.7	580	18	US-10-021-323-12291	Sequence 12291, A

ALIGNMENTS

RESULT 1
US-10-042-059A-1

Sequence 1, Application US/10042059A
; Publication No. US20020197704A1
;
; GENERAL INFORMATION:
; APPLICANT: Rhein Biotech Gesellschaft fur neue biotechnologische Prozesse und
; APPLICANT: Produkte mbH
; TITLE OF INVENTION: Nucleic Acid Molecule Containing a Nucleic Acid Coding for a Pol
; TITLE OF INVENTION: with Chorismate Mutase Activity
; FILE REFERENCE: PA3058US-019
; CURRENT APPLICATION NUMBER: US/10/042,059A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: DE 199 19 124.7
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Hansenula polymorpha
US-10-042-059A-1

Query Match   100.0%; Score 843; DB 13; Length 843;
Best Local Similarity 100.0%;
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGGACTTTATGAAGCCAGAAAACAGTGCTGGACCTTGGCAACATTAGAGATGCCTTGGTC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACTTTATGAAGCCAGAAAACAGTGCTGGACCTTGGCAACATTAGAGATGCCTTGGTC 60

Qy   61 CGGATGGGAGGATACGGATCATCTTCAACTTTATCGGAGCGGGTCGCAGTTCTATGCGTCC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CGGATGGGAGGATACGGATCATCTTCAACTTTATCGGAGCGGGTCGCAGTTCTATGCGTCC 120

```
                              GenCore version 5.1.6
                      Copyright (c) 1993 - 2005 Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:         March 31, 2005, 00:02:41 ; Search time 3559 Seconds
                                            (without alignments)
                                            9016.066 Million cell updates/sec

Title:          US-10-042-059B-1
Perfect score:  843
Sequence:       1 atggacttcatgaagccaga..........tctcctcctggcttgtactag 843

Scoring table:  IDENTITY_NUC
                Gapop:10.0 , Gapext 1.0

Searched:       34239544 segs, 1903213470 residues

Total number of hits satisfying chosen parameters:   68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_est7:*
                9:  gb_gss2:*

Pred. No. is,the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                  SUMMARIES

Result                       %
No.     Score    Query    Match  DB  ID             Description
                 Length
-----   -----    ------   -----  --  --             -----------
    1   299      35.5     1185    9  CNS06T5P       AL414131 T7 end of
c   2   176.6    20.9     605     7  CNS06TV7       AL415049 T3 end of
c   3   176.6    20.9     1052    9  CNS06XEJ       AL419633 T3 end of
c   4   140      16.6     948     7  CNS06472       CO005472 T3 end of
    5   137      16.3     574     8  CNS06472       EST793807
    6   132.8    15.8     738     7  CF877247       KD2284.p1
    7   132.8    15.8     816     6  CB905545       tric075xb
c   8   125.4    14.9     788     7  CO165801       tric075xb
c   9   121.6    14.4     660     5  BX253798       FLD1_57_E
   10   121.6    13.9     798     6  BX253798       BX253798
c  11   116.2    13.8     652     6  CB619881       OSIIEa04L
c  12   113.8    13.5     747     6  CD423256       SA1_28_B1
c  13   113.8    13.4     710     7  CNS06TWI       T3_end_of
   14   111.2    13.2     1222    3  AY103806       RTWW2_11_
   15   107.4    12.7     827     1  AY103806       Zea_mays
   16   106      12.6     988     1  CO162312       FLD1_34_B
   17   105.8    12.6     602     6  CD982471       CD982471
   18   105.8    12.5     903     7  CN149106       OAK5c04.x
c  19   105.6    12.5     527     7  CN149106       WOUND1_60
c  20   104.6    12.4     606     4  BI995968       FLD1_57_F
c  21   103      12.2     737     7  CO165812       103103260
c  22   103      12.2     867     6  CB619882       EST793808
c  23   102.8    12.2     727     5  CA121888       OSIIBa04L
   24   102.2    12.1     680     6  CA080410       SCVPAM105
```

```
RESULT 1
CNS06T5P
LOCUS          CNS06T5P                  1185 bp    DNA    linear    GSS 05-JUL-2001
DEFINITION     T7 end of clone AW0AA029A11 of library AW0AA from strain CLIB 89 of
               Yarrowia lipolytica, genomic survey sequence.
ACCESSION      AL414131
VERSION        AL414131.1  GI:12186949
KEYWORDS       GSS.
SOURCE         Yarrowia lipolytica
  ORGANISM     Yarrowia lipolytica
               Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE      1  (bases 1 to 1185)
  AUTHORS      Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
               Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
               de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
               Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
               Saurin,W., Tekaia,F., Toffano-Nioche,C., Wésolowski-Louvel,M.,
               Wincker,P. and Weissenbach,J.
  TITLE        Genomic exploration of the hemiascomycetous yeasts: 1. A set of
               yeast species for molecular evolution studies
  JOURNAL      FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE      20584711
   PUBMED      11115876
REFERENCE      2  (bases 1 to 1185)
  AUTHORS      Casaregola,S., Neuvéglise,C., Lepingle,A., Bon,E., Feynerol,C.,
               Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE        Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
               lipolytica
  JOURNAL      FEBS Lett. 487 (1), 95-100 (2000)
  MEDLINE      20584727
   PUBMED      11152892
REFERENCE      3  (bases 1 to 1185)
  AUTHORS      Genoscope.
  TITLE        Direct Submission
  JOURNAL      Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
               2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE, (E-mail :
               seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT        This GSS is part of a random genomic sequencing program of thirteen
               yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
               exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
               Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
               lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
               angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
               Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
               5 kb were prepared and both extremities were sequenced. See
               keywords for description of this sequence and for the sequence of
```

```
                                 ALIGNMENTS

    25   100.8    12.0     551     5  BQ701847       BQ701847  NXSI_121
    26   99.4     11.8     605     7  CN913779       CN913779  030108ABM
    27   98.6     11.7     607     8  BZ296833       BZ296833  CG3070.r1
    28   98       11.6     495     1  AA786587       AA786587  mb508a1.f
c   29   97.8     11.6     821     7  CK196770       CK196770  FGA500623
    30   97.4     11.6     837     7  CO366854       CO366854  RTK1_30_C
    31   97       11.5     596     8  BZ298767       BZ298767  CG4569.r1
    32   97       11.5     781     7  CF667838       CF667838  RTCNT1_32
c   33   96.6     11.5     853     7  CO366784       CO366784  RTK1_30_C
    34   96.2     11.4     657     9  AG273371       AG273371  Cyanidios
c   35   96       11.4     812     7  CN146636       CN146636  WOUND1_57
    36   95       11.3     885     7  CV274873       CV274873  WS0174.B2
c   37   94.8     11.2     857     7  CF635400       CF635400  zmrww00.0
c   38   94.6     11.2     480     7  CV002646       CV002646  nad03-13c
    39   94.4     11.2     707     6  CA239381       CA239381  SCRFFL503
    40   94.2     11.2     785     7  CNS23137       CNS23137  GQ015M01.
    41   94.2     11.1     552     4  BI718858       BI718858  10310320
c   42   93.6     11.1     1265    3  CNS06WL        CNS0A727   Arabidops
    43   93.6     11.1     1329    3  BX822129       BX824460   Arabidops
    44   93.2     11.1     698     2  BE577489       BE577489  L48-2335T
c   45   93.2     11.1     830     6  CB625596       CB625596  OSIIEa09I
```

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 08:01:45 ; Search time 88 Seconds
(without alignments)
1230.601 Million cell updates/sec

Title: 1434
Perfect score: 1434
Sequence: 1 MDFMKPETVLDLGNIRDALV.........DDATQKSGGYVDRPLSSGLY 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_16Dec04:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003s:*
8: geneseqp2003as:*
8: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1434	100.0	280	3	AAB11451	Aab11451 H. polymo
2	921	64.2	268	8	ADP98857	Adp98857 C. albica
3	587.5	41.0	266	8	ADR86097	Adr86097 Aspergill
4	581.5	40.6	263	6	AAO19566	Aao19566 M sterili
5	508.5	35.5	257	8	ABU08093	Abu08093 Rice chor
6	508.5	35.5	257	8	ADJ55264	Adj55264 Soybean c
7	481.5	33.6	316	8	ABB91515	Abb91515 Herbicida
8	481.5	33.6	331	8	ABU08092	Abu08092 Corn chor
9	481.5	33.6	331	8	ADJ55262	Adj55262 Corn chor
10	466	32.5	334	5	ABB92474	Abb92474 Herbicida
11	464	32.4	334	8	ABU08097	Abu08097 A. thalia
12	464	32.4	334	8	ADJ55271	Adj55271 Arabidops
13	423	29.5	261	6	ABU08094	Abu08094 Soybean c
14	423	29.5	261	8	ADJ55266	Adj55266 Soybean c
15	399	27.8	265	5	AAM49654	Aam49654 A. thalia
16	399	27.8	265	8	ABB93424	Abb93424 Herbicida
17	399	27.8	265	8	ABU08098	Abu08098 A. thalia
18	399	27.8	265	8	ADJ55272	Adj55272 Arabidops
19	399	27.8	330	5	AAM49655	Aam49655 Construct
20	391	27.3	319	3	AAG06003	Aag06003 Arabidops
21	391	27.3	325	3	AAG06002	Aag06002 Arabidops
22	391	27.3	347	3	AAG06001	Aag06001 Arabidops
23	313	21.8	207	6	ABU08095	Abu08095 Wheat cho
24	313	21.8	207	6	ADJ55268	Adj55268 Wheat cho
25	221	15.4	154	6	ABU08090	Abu08090 Protein e
26	221	15.4	154	8	ADJ55258	Adj55258 Wheat cho
27	197.5	13.8	132	6	ABU08089	Abu08089 Protein e
28	197.5	13.8	132	8	ADJ55256	Adj55256 Soybean c
29	132	9.2	93	8	ABU08087	Abu08087 Protein e
30	132	9.2	93	8	ADJ55252	Adj55252 Corn chor
31	122	8.5	62	8	ABU08088	Abu08088 Protein e
32	122	8.5	62	8	ADJ55254	Adj55254 Rice chor
33	114.5	8.0	469	2	AAY20059	Aay20059 B. burgdo
34	114.5	8.0	446	2	AAY20058	Aay20058 B. burgdo
35	105.5	7.4	1638	9	ADK71824	Adk71824 Human kin
36	105.5	7.2	1732	8	ADF95103	Adf95103 Rat serin
37	103.5	7.2	1732	8	ABG15384	Abg15384 Novel hum
38	103	7.2	724	4	ABG17531	Abg17531 Novel hum
39	103	7.2	552	6	ABR43675	Abr43675 Sarcina v
40	102	7.1	1664	7	ADE47740	Ade47740 Human NOV
41	100.5	7.0	1664	8	ADJ79010	Adj79010 Human NOV
42	100.5	7.0	1719	8	AAE21707	Aae21707 Human PKI
43	100.5	7.0	1732	8	ADF89991	Adf89991 Human ser
44	100.5	7.0	1732	8	ADS93536	Ads93536 Human MRC
45	100.5	7.0	1732	8		

ALIGNMENTS

RESULT 1
AAB11451

ID AAB11451 standard; protein; 280 AA.
XX
AC AAB11451;
XX
DT 12-SEP-2003 (revised)
DT 01-MAR-2001 (first entry)
XX
DE H. polymorpha chorismate mutase protein.
XX
KW Chorismate mutase; prephenate; selection marker; auxotrophic yeast.
XX
OS Pichia angusta.
XX
PN WO200065071-A1.
XX
PD 02-NOV-2000.
XX
PF 27-APR-2000; 2000WO-EP003844.
XX
PR 27-APR-1999; 99DE-01019124.
XX
PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX
PI Gellissen G, Braus G, Pries R, Krappmann S, Strasser AW;
XX
DR WPI; 2000-687355/67.
DR N-PSDB; AAC81949.
XX
PT Nucleic acids encoding chorismate mutase, useful for preparing an
PT auxotrophic selection system for recombinant yeast and for recombinant
PT protein expression.
XX
PS Claim 1h; Page 57-58; 63pp; German.
XX
CC This invention describes novel nucleic acids (I) that encode a
CC polypeptide (II) with chorismate mutase (CM) activity (or its
CC complementary strand), CM catalyzes conversion of chorismate to
CC prephenate, an essential precursor for Phe and Tyr. (I) is a selection
CC marker for construction of corresponding auxotrophic yeast (requiring Phe
CC and Tyr) that are used for recombinant production of proteins. (I) allows
CC selection of transformed yeast on simple media. (Updated on 12-SEP-2003
CC to standardise OS field)
XX
SQ Sequence 280 AA;

Query Match 100.0%; Score 1434; DB 3; Length 280;

```
OM protein - protein search, using sw model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:             March 31, 2005, 10:14:39 ; Search time 29 Seconds
                                              (without alignments)
                                              720.750 Million cell updates/sec

Title:              US-10-042-059B-2
Perfect score:      1434
Sequence:           1 MDFMKPETVLDLGNIRDALV.........DDATQKSGGYVDRFLSSGLY 280

Scoring table:      BLOSUM62
                    Gapop 10.0 , Gapext 0.5
Searched:           513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters:    513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        Issued Patents AA:*
                  1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                  2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                  3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                  4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                  5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                  6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                SUMMARIES

Result                      Query    Match
No.     Score   %           Length   DB      ID                      Description
-----   -----   ---------   ------   ---     --------------------    -----------
1       921     64.2        315      4       US-09-248-796A-18213    Sequence 18213, A
2       508.5   35.5        257      4       US-09-454-279-14        Sequence 14, Appl
3       481.5   33.6        331      4       US-09-454-279-12        Sequence 12, Appl
4       464     32.4        334      4       US-09-454-279-21        Sequence 21, Appl
5       423     29.5        261      4       US-09-454-279-16        Sequence 16, Appl
6       399     27.8        265      4       US-09-454-279-22        Sequence 22, Appl
7       313     21.8        207      4       US-09-454-279-18        Sequence 18, Appl
8       221     15.4        154      4       US-09-454-279-8         Sequence 8, Appl
9       197.5   13.8        132      4       US-09-454-279-6         Sequence 6, Appl
10      132     9.2         132      4       US-09-454-279-2         Sequence 2, Appl
11      122     8.5         62       4       US-09-454-279-4         Sequence 4, Appl
12      94      6.6         391      4       US-09-270-767-42309     Sequence 42309, A
13      94      6.6         502      4       US-09-635-872A-13       Sequence 13, Appl
14      94      6.6         502      4       US-09-636-077A-13       Sequence 13, Appl
15      94      6.6         502      4       US-09-636-060C-13       Sequence 13, Appl
16      94      6.6         502      4       US-09-636-552-13        Sequence 13, Appl
17      94      6.6         502      4       US-09-986-552-13        Sequence 13, Appl
18      94      6.6         502      4       US-09-636-596C-13       Sequence 13, Appl
19      94      6.6         502      4       US-10-023-894-16        Sequence 16, Appl
20      92      6.4         717      4       US-10-306-686-13        Sequence 13, App
21      92      6.4         721      4       US-09-583-101-4629      Sequence 4629, Ap
22      90      6.3         292      4       US-09-107-433-2819      Sequence 2819, Ap
23      89.5    6.2         640      4       US-09-107-532A-3741     Sequence 3741, Ap
24      89      6.2         335      4       US-09-248-796A-16129    Sequence 16129, A
25      89      6.2         335      4       US-09-934-901-6         Sequence 6, Appl
26      89      6.2         335      4       US-09-934-868-16        Sequence 16, Appl
27      89      6.2         335      4       US-10-320-874-6         Sequence 6, Appli
28      88      6.1         315      4       US-09-248-796A-18213    Sequence 18213, A
29      87.5    6.1         257      4       US-09-106-872A-14616    Sequence 14616, A
30      87      6.1         626      4       US-08-694-869-1         Sequence 1, Appli
31      87      6.1         1871     2       US-08-349-546-1         Sequence 1, Appli
32      87      6.1         1871     3       US-09-502-831-1         Sequence 1, Appli
33      85      5.9         883      4       US-08-982-430-1         Sequence 1, Appli
34      85      5.9         1165     2       US-09-640-389A-11       Sequence 11, Appl
35      85      5.9         1165     3       US-09-093-814-1         Sequence 1, Appli
36      85      5.9         1165     3       US-08-618-957A-11       Sequence 11, Appl
37      85      5.9         1165     4       US-10-095-929-11        Sequence 11, Appl
38      85      5.9         1221     4       US-09-248-796A-19541    Sequence 19541, A
39      84.5    5.9         234      4       US-09-710-279-2782      Sequence 2782, Ap
40      84.5    5.9         284      4       US-09-710-279-482       Sequence 482, App
41      84.5    5.9         610      4       US-09-134-001C-3061     Sequence 3061, Ap
42      84.5    5.9         614      4       US-09-710-279-2464      Sequence 2464, Ap
43      84.5    5.9         625      3       US-09-710-279-2464      Sequence 2464, Ap
44      84.5    5.9         631      3       US-09-134-001C-4605     Sequence 4605, Ap
45      84.5    5.9         661      4       US-09-710-279-274       Sequence 274, App
```

```
                                ALIGNMENTS

RESULT 1
US-09-248-796A-18213
; Sequence 18213, Application US/0924879GA
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1999-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
;  SEQ ID NO 18213
;  LENGTH: 315
;  TYPE: PRT
;  ORGANISM: Candida albicans
US-09-248-796A-18213

  Query Match             64.2%;  Score 921;  DB 4;  Length 315;
  Best Local Similarity   65.6%;  Pred. No. 4.7e-88;
  Matches    170;  Conservative  44;  Mismatches  45;  Indels  0;  Gaps  0;

Qy     1 MDFMKPETVLDLGNIRDALVRMEDTIIFNFIERSQFYASPSVYKNQFPIPNFDGSFLDW  60
           |||||||||||||||||||| |||:| |:|||||||||:|||||||||:||||||:|:|
Db    48 MDFMKPETVLDLANIRQALVRMEDTIVFDLIERSQFFSSPSVYEKNKYNIPNFDGTFLEW  107

Qy    61 LLSQHERIIHSQVRRYDAPDEVPFPNVLEKTFLPKINYPSVLASYADEINVNKEILKIYT  120
           |||:|||||:||::|:||||||||||||:|||||:||||||:|||:|||||||||||||
Db   108 ALLQLEVAHSQIRYEAPDETPFFPDQLKTPILPPINYPKILAKYSDEINVNSEIMKFYV  167

Qy   121 SEIVPGIAAGSGEQEDNLGSCAMADIECLLQSLSRRIHFGRPVAEAKPISEGDKIVDLIKK  180
           |:|||:||||||||||:||||:||||:|||:|||:|||||||||:|:||:|||||||:||
Db   168 DEIVPQVSCCQGDQKENLGSASTCDIECLQAISRRIHFGKPVAEAKYQSDKPLYIKLLD  227

Qy   181 RDVEGIEALITNAEVEKRILDRLEKGRAYGTDPTLKFTQHIQSKYKPEVIVKIYKDFVI  240
           :|||:|||| |:||||||||||||||||||||:||||||||||||||||||:||||||
Db   228 KDVKGIENSITNSAVEQKILERLIVKAESYGVDPSLKFGQNVQSKVKPEVIAKLYKDWII  287

Qy   241 PLTKVEVDYLLRRLEDEH  259
           |||||:||||||||||||:
Db   288 PLTKKVEIDYLLRRLEDED  306

RESULT 2
US-09-454-279-14
; Sequence 14, Application US/0945427Y
```

```
OM protein - protein search, using sw model

              GenCore version 5.1.6
              Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:       March 31, 2005, 10:17:14 ; Search time 75 Seconds
                                          (without alignments)
                                          1237.947 Million cell updates/sec

Title:        US-10-042-059B-2
Perfect score: 1434
Sequence:     1 MDFMKPETVLDLGNIRDALV.........DDATQKSGGYVDRFLSSGLY 280

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5
Searched:     1413372 seqs, 331592847 residues

Total number of hits satisfying chosen parameters:    1413372

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Published Applications AA:*
          1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
          2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
          3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
          4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
          5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
          6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
          7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
          8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
          9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
         10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
         11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
         12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
         13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
         14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
         15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
         16:  /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
         17:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
         18:  /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
         19:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
         20:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

SUMMARIES

```
Result              %
No.   Score  Query Match Length DB ID        Description
----  -----  ----------- ------ -----        -----------
1     1434   100.0    280   13  US-10-042-059A-2  Sequence 2, Appli
2     921    64.2     268   17  US-10-741-849-7032  Sequence 7032, Ap
3     508.5  35.5     257    9  US-09-454-279-14  Sequence 14, Appl
4     508.5  35.5     257   15  US-10-624-061-14  Sequence 14, Appl
5     508.5  35.5     313   16  US-10-437-963-196439  Sequence 196439,
5     497.5  34.7     320   15  US-10-425-114-68586  Sequence 68586, A
7     495.5  34.6     315   15  US-10-424-599-266662  Sequence 266662,
8     486    33.9     350   16  US-10-767-701-45917  Sequence 45917, A
9     481.5  33.6     331   16  US-09-454-279-12  Sequence 12, Appl
10    481.5  33.6     331   15  US-10-624-061-12  Sequence 12, Appl
11    464    32.4     334    9  US-09-454-279-21  Sequence 21, Appl
12    464    32.4     334   15  US-10-624-061-21  Sequence 21, Appl
13    423    29.5     261    9  US-10-454-279-16  Sequence 16, Appl
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
14    423    29.5     261   15  US-10-624-061-16  Sequence 16, Appl
15    423    29.5     297   15  US-10-425-114-48703  Sequence 48703, A
16    419    29.2     261   15  US-10-424-599-235878  Sequence 235878,
17    408.5  28.5     255   16  US-10-437-963-151855  Sequence 151855,
18    399    27.8     265    9  US-09-454-279-22  Sequence 22, Appl
19    399    27.8     265   15  US-10-624-061-22  Sequence 22, Appl
20    374.5  26.1     313   15  US-10-425-114-65608  Sequence 65608, A
21    373    26.0     264   15  US-10-425-114-66528  Sequence 66528, A
22    356    24.8     306   16  US-10-437-963-186047  Sequence 186047,
23    348    24.3     316   16  US-10-437-963-186040  Sequence 186040,
24    344.5  24.0     186   15  US-10-425-114-52677  Sequence 52677, A
25    328    22.9     199   15  US-10-425-114-43606  Sequence 43606, A
26    313    21.8     207    9  US-09-454-279-18  Sequence 18, Appl
27    313    21.8     207   15  US-10-624-061-18  Sequence 18, Appl
28    296.5  20.7     489   16  US-10-437-963-103185  Sequence 103185,
29    282    19.7     210   16  US-10-437-963-103185  Sequence 32934, A
30    221    15.4     154    9  US-09-454-279-8   Sequence 8, Appli
31    221    15.4     154   15  US-10-624-061-8   Sequence 8, Appli
32    197.5  13.8     132   15  US-10-624-061-6   Sequence 6, Appli
33    197.5  13.8     132    9  US-09-454-279-6   Sequence 6, Appli
34    178    12.4     109   15  US-10-424-599-224668  Sequence 224668,
35    157.5  11.0      97   15  US-10-425-114-38176  Sequence 38176, A
36    132     9.2      93   15  US-09-454-279-2   Sequence 2, Appli
37    132     9.2      93   15  US-10-624-061-2   Sequence 2, Appli
38    122     8.5      62    9  US-09-454-279-4   Sequence 4, Appli
39    122     8.5      62   15  US-10-624-061-4   Sequence 4, Appli
40    106.5   7.4      89   15  US-10-424-599-272410  Sequence 272410,
41    104.5   7.3      83   15  US-10-425-114-52545  Sequence 52545, A
42    102     7.1     552   14  US-10-136-960-6   Sequence 6, Appli
43    100.5   7.0     458   15  US-09-771-161A-137  Sequence 137, Appl
44    100.5   7.0    1664   15  US-10-210-130-102  Sequence 102, App
45    100.5   7.0    1719   15  US-10-288-798-2   Sequence 2, Appli
```

ALIGNMENTS

RESULT 1
US-10-042-059A-2
; Sequence 2, Application US/10042059A
; Publication No. US2002019770A1
; GENERAL INFORMATION:
; APPLICANT: Rhein Biotech Gesellschaft fur neue biotechnologische Prozesse und
; APPLICANT: Produkte mbH
; TITLE OF INVENTION: Nucleic Acid Molecule Containing a Nucleic Acid Coding for a Pol
; TITLE OF INVENTION: with Chorismate Mutase Activity
; FILE REFERENCE: PA3055BUS-019
; CURRENT APPLICATION NUMBER: US/10/042,059A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: DE 199 19 124.7
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Hansenula polymorpha
US-10-042-059A-2

Query Match   100.0%; Score 1434; DB 13; Length 280;
Best Local Similarity   100.0%;
Matches 280; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MDFMKPETVLDLGNIRDALVRMEDIIFNFIERSQFYASPSVYKVNQFPIPNFDGSFLDW 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MDFMKPETVLDLGNIRDALVRMEDIIFNFIERSQFYASPSVYKVNQFPIPNFDGSFLDW 60

Qy   61 LLSQHERIHSQVRRYDAPDEVFPFFPNVLEKTFLPKINYPSVLASYADEINVNKEILKIYT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LLSQHERIHSQVRRYDAPDEVFPFFPNVLEKTFLPKINYPSVLASYADEINVNKEILKIYT 120

Qy   121 SEIVPGIAAGSGEQEDNLGSCAMADIECLQSLSRRIHFGRFVAEAKFISEGDKIVDLIKK 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 SEIVPGIAAGSGEQEDNLGSCAMADIECLQSLSRRIHFGRFVAEAKFISEGDKIVDLIKK 180
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on:          March 31, 2005, 10:10:35 ; Search time 26' Seconds
                 (without alignments)
                 1036.181 Million cell updates/sec

Title:           US-10-042-059B-2
Perfect score:   1434
Sequence:        1 MDFMKPETVLDIGNIRDALV.........DDATQKSGGYVDRFLSSGLY 280

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:          283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_79:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 776.5 | 54.1 | 256 | 2 | A45921 | chorismate mutase |
| 2 | 558.5 | 38.9 | 251 | 2 | T37784 | probable chorismat |
| 3 | 464 | 32.4 | 334 | 2 | S38958 | chorismate mutase |
| 4 | 399 | 27.8 | 265 | 2 | T50796 | chorismate mutase |
| 5 | 166 | 11.6 | 82 | 2 | T14902 | chorismate mutase |
| 6 | 119 | 8.3 | 70 | 2 | T14901 | chorismate mutase |
| 7 | 114.5 | 8.0 | 469 | 2 | C70109 | hypothetical prote |
| 8 | 104 | 7.3 | 313 | 2 | B75138 | tetrahydropteroylt |
| 9 | 104 | 7.3 | 336 | 2 | S04739 | site-specific DNA- |
| 10 | 103.5 | 7.2 | 1732 | 2 | T14039 | protein kinase (EC |
| 11 | 99 | 6.9 | 1033 | 2 | T37715 | actin-interacting |
| 12 | 98 | 6.8 | 1702 | 2 | T14050 | protein kinase (EC |
| 13 | 97.5 | 6.8 | 1286 | 2 | T37366 | RNA polymerase sub |
| 14 | 97 | 6.8 | 323 | 2 | D90527 | glycerol-3-phospha |
| 15 | 97 | 6.8 | 1167 | 2 | E96963 | DNA polymerase III |
| 16 | 96.5 | 6.7 | 829 | 2 | S72366 | DNA topoisomerase |
| 17 | 96.5 | 6.7 | 932 | 2 | F84465 | hypothetical prote |
| 18 | 96 | 6.7 | 808 | 2 | H64474 | hypothetical prote |
| 19 | 96 | 6.7 | 1286 | 2 | H36845 | DNA-directed RNA p |
| 20 | 96 | 6.7 | 1286 | 2 | T28521 | DNA-directed RNA p |
| 21 | 96 | 6.7 | 1286 | 2 | A71161 | M6R protein - vari |
| 22 | 95.5 | 6.7 | 560 | 2 | F69059 | arginine-tRNA liga |
| 23 | 94.5 | 6.6 | 790 | 2 | T19683 | hypothetical prote |
| 24 | 94.5 | 6.6 | 1287 | 2 | RNV247 | DNA-directed RNA p |
| 25 | 94.5 | 6.6 | 7829 | 1 | T15789 | hypothetical prote |
| 26 | 94 | 6.6 | 311 | 2 | C69952 | probable ribosomal |
| 27 | 94 | 6.6 | 502 | 2 | AF1310 | probable thermosta |
| 28 | 94 | 6.6 | 612 | 2 | S73611 | glucose inhibited |
| 29 | 93 | 6.5 | 307 | 2 | A75020 | hypothetical prote |
| 30 | 92 | 6.4 | 992 | 2 | T39915 | hypothetical prote |
| 31 | 91.5 | 6.4 | 289 | 2 | D70440 | DNA polymerase I 3 |
| 32 | 91.5 | 6.4 | 456 | 2 | G81408 | probable outer mem |
| 33 | 91.5 | 6.4 | 474 | 2 | S38833 | glutathione syntha |
| 34 | 91.5 | 6.4 | 663 | 2 | AF1799 | hypothetical prote |
| 35 | 91.5 | 6.4 | 913 | 2 | T15474 | hypothetical prote |
| 36 | 91 | 6.3 | 345 | 2 | C81300 | probable helix-tur |
| 37 | 91 | 6.3 | 377 | 2 | B90437 | glycosyltransferas |
| 38 | 91 | 6.3 | 398 | 2 | B97341 | hypothetical prote |
| 39 | 91 | 6.3 | 551 | 2 | B90247 | DNA repair protein |
| 40 | 91 | 6.3 | 596 | 2 | G82874 | conserved hypothet |
| 41 | 90.5 | 6.3 | 475 | 2 | T33979 | hypothetical prote |
| 42 | 90 | 6.3 | 431 | 2 | H71172 | hypothetical prote |
| 43 | 90 | 6.3 | 445 | 2 | T23119 | hypothetical prote |
| 44 | 89.5 | 6.2 | 583 | 2 | C90327 | n-methylhydantoina |
| 45 | 89.5 | 6.2 | 859 | 2 | T06352 | lipoxygenase (EC 1 |

ALIGNMENTS

RESULT 1
A45921
chorismate mutase (EC 5.4.99.5) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YP9499.15c; protein YPR060c
C;Species: Saccharomyces cerevisiae
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C;Accession: A45921; S54081
R;Schmidheini, T.; Sperisen, P.; Paravicini, G.; Huetter, R.; Braus, G.
J. Bacteriol. 171, 1245-1253, 1989
A;Cross-references: UNIPROT:P32178; GB:M24517; NID:g295576; PIDN:AAB59309.1; PID:g2955?
A;Title: A single point mutation results in a constitutively activated and feedback-re
R;Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54059
A;Accession: S54081
A;Molecule type: DNA
A;Residues: 1-256 <SCH>
A;Cross-references: A45921; S54081
A;Molecule type: DNA
A;Accession: A45921
A;Map position: 16R
C;Function:
A;Description: intramolecular transferase; isomerase
A;Description: chorismate mutase
C;Genetics:
A;Gene: SGD;ARO7
A;Cross-references: SGD:S0006264; MIPS:YPR060c
C;Genetics:
A;Residues: 1-256 <BAD>
A;Cross-references: EMBL:Z49219; NID:g805025; PIDN:CAA89177.1; PID:g805040; MIPS:YPR06C
A;Experimental source: strain AB972
A;Pathway: aromatic amino acid biosynthesis
C;Superfamily: chorismate mutase of the AroQ class, eukaryotic type
C;Keywords: aromatic amino acid biosynthesis; intramolecular transferase; isomerase

Query Match      54.1%;     Score 776.5;  DB 2;  Length 256;
Best Local Similarity   58.7%;  Pred. No. 2.1e-50;
Matches  152;   Conservative  40;  Mismatches  62;  Indels  5;  Gaps  2;

Qy         1  MDFMKPETVLDIGNIRDALVRMEDTIIFNFIERSQFYASPSVTKVNQ--FPIPNFDGSFL  58
              ||||||||||||||||||||||:||:|||||||||||||||||||||  ||:||||||||
Db         1  MDFIKPETVLNLQNIRDELVRMEDSIIFKFIERSHFATCPSVYEANHPGLEIPNFKGSFL  60

Qy        59  DWLLSQHERIHSQVRRYDAPDEVPFPNVLEKTFLPKINYPSVLASYADEINVKBILKI  118
              || |||:|||||:|||||||:: :|:|||:||:::|||||||:||:||:|||:||:|:
Db        61  DWALSNLBIAHSRIRRFRFSPDETFPFPDKIQKSFLPSINYPQILAPYAPEVNYNDKIKKV  120

Qy       119  YTSEIVPGIAAGSGBQEDENLGSCAMADIECLQSLSRRIHFGRFVAEAKFISEGDKIVDLI  178
              ||:||||||||||||||:::::||||||:|||||:||||:||||||:|:||||||:||||:
Db       121  YIEKIIPLISKRDGDDKNNFGSVATRDIECLQSLSRRIHFGKFVAEAKFQSDIPLYTKLI  180

Qy       179  KKRDVEGIEALITNAEVEKRILDRLLEKGRAYGTDPTLKFTQHIQSKVAPEVIVKIYKDF  238
              |:||||||||||::|||||||||||||||||||||||||||||||||||||||||||||
Db       181  KSKDVEGIMKNITNSAVEEKILERLITKKAEVGVDPT---NESGERRIPEYLVKIYKEI  237

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:        March 31, 2005, 09:04:29 ; Search time 85 seconds
                                          (without alignments)
                                          1686.849 Million cell updates/sec

Title:         US-10-042-059B-2
Perfect Score: 1434
Sequence:      1 MDFMKPETVLDLGNIRDALV.........DDATQKSGGYVDRFLSSGIY 280

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters:        1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     UniProt_03:*
               1: uniprot_sprot:*
               2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 1434 | 100.0 | 280 | 2 | Q9P4D8 | pichia angu |
| 2 | 907 | 63.2 | 267 | 2 | Q6BK64 | debaryomyce |
| 3 | 808 | 56.3 | 256 | 2 | Q6CSJ7 | yarrowia li |
| 4 | 776.5 | 54.1 | 256 | 1 | CHMU_YEAST | saccharomyc |
| 5 | 752.5 | 52.5 | 260 | 2 | Q6CVY3 | kluyveromyc |
| 6 | 719 | 50.1 | 260 | 2 | Q6FLZ7 | candida gla |
| 7 | 702.5 | 49.0 | 259 | 2 | Q75BG5 | ashbya goss |
| 8 | 594 | 41.4 | 269 | 2 | Q75BR4 | neurospora |
| 9 | 581.5 | 40.6 | 263 | 2 | Q618Q0 | roselinia |
| 10 | 579 | 40.4 | 267 | 2 | Q6Y7B2 | emericella |
| 11 | 558.5 | 38.9 | 251 | 1 | CHMU_SCHPO | schizosacch |
| 12 | 508.5 | 35.5 | 294 | 2 | Q93Y60 | oryza sativ |
| 13 | 481.5 | 33.6 | 316 | 2 | Q9C544 | arabidopsis |
| 14 | 475 | 33.1 | 295 | 2 | Q96V28 | cryptococcu |
| 15 | 466 | 32.5 | 334 | 1 | CHMU_ARATH | arabidopsis |
| 16 | 465 | 32.4 | 255 | 2 | Q9STE2 | lycopersico |
| 17 | 464.5 | 32.4 | 316 | 2 | Q9XF60 | arabidopsis |
| 18 | 408.5 | 28.5 | 255 | 2 | Q6H819 | oryza sativ |
| 19 | 399 | 27.8 | 265 | 2 | Q9S7H4 | arabidopsis |
| 20 | 391 | 27.3 | 284 | 2 | Q62BE6 | oryza sativ |
| 21 | 290.5 | 20.3 | 179 | 2 | Q6Z9E6 | oryza sativ |
| 22 | 247 | 17.2 | 131 | 2 | Q6F3R1 | nicotiana t |
| 23 | 166 | 11.6 | 82 | 2 | Q65861 | prunus aviu |
| 24 | 119 | 8.3 | 70 | 2 | O22410 | petroselinu |
| 25 | 114.5 | 8.0 | 469 | 2 | O51102 | petroselinu |
| 26 | 111 | 7.7 | 637 | 2 | Q65IH3 | bacillus li |
| 27 | 107 | 7.5 | 622 | 2 | Q7NAK6 | mycoplasma |
| 28 | 106.5 | 7.4 | 413 | 2 | Q8R3P7 | mus musculu |
| 29 | 106 | 7.4 | 1286 | 2 | Q8V4Z4 | monkeypox v |
| 30 | 106 | 7.4 | 5289 | 2 | Q8BZ03 | lactobacill |
| 31 | 105.5 | 7.4 | 1638 | 2 | Q81WQ7 | homo sapien |
| 32 | 104 | 7.3 | 313 | 2 | Q9Y086 | pyrococcus |
| 33 | 104 | 7.2 | 336 | 1 | MTP2_PROVU | proteus vul |
| 34 | 103.5 | 7.2 | 1732 | 2 | O54874 | rattus norv |
| 35 | 102.5 | 7.1 | 611 | 2 | Q9KX66 | mycoplasma |
| 36 | 102 | 7.1 | 311 | 2 | Q6SH56 | bacillus li |
| 37 | 102 | 7.1 | 345 | 2 | P70904 | borrelia he |
| 38 | 102 | 7.0 | 552 | 2 | Q93EN4 | sarcina ven |
| 39 | 100.5 | 7.0 | 357 | 2 | Q8V0L9 | staphylococ |
| 40 | 100.5 | 7.0 | 355 | 2 | Q8VY0 | staphylococ |
| 41 | 100.5 | 7.0 | 1638 | 2 | Q86XX3 | staphylococ |
| 42 | 100.5 | 7.0 | 1719 | 2 | Q86XX2 | homo sapien |
| 43 | 100 | 6.9 | 450 | 2 | Q74K27 | homo sapien |
| 44 | 99.5 | 6.9 | 492 | 1 | K2CO_CHICK | gallus gall |
| 45 | 99 | 6.9 | 352 | 2 | Q74MI4 | nanoarchaeu |

ALIGNMENTS

RESULT 1
Q9P4D8

ID   Q9P4D8          PRELIMINARY;     PRT;   280 AA.
AC   Q9P4D8;
DT   01-OCT-2000 (TrEMBlrel. 15, Created)
DT   01-OCT-2000 (TrEMBlrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBlrel. 26, Last annotation update)
DE   Chorismate mutase (EC 5.4.99.5).
GN   Name=HARO7;
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4905;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB11;
RX   MEDLINE=20353451; PubMed=10894726;
RX   DOI=10.1128/JB.182.15.4188-4197.2000;
RA   Krappmann S., Pries R., Gellissen G., Hiller M., Braus G.H.;
RT   "HARO7 encodes chorismate mutase of the methylotrophic yeast Hansenula
RT   polymorpha and is derepressed upon methanol utilization.";
RL   J. Bacteriol. 182:4188-4197(2000).
DR   EMBL; AF204738; AAF87954.1; -.
DR   HSSP; P32178; 2CSM.
DR   GO; GO:0004106; F:chorismate mutase activity; IEA.
DR   GO; GO:0016853; F:isomerase activity; IEA.
DR   GO; GO:0009073; P:aromatic amino acid family biosynthesis; IEA.
DR   InterPro; IPR002701; Chorismate_mut.
DR   InterPro; IPR008238; Chor_mut_AroQ_eu.
DR   Pfam; PF01817; CM_2; 1.
DR   PIRSF; PIRSF017318; Chor_mut_AroQ_eu; 1.
DR   TIGRFAMs; TIGR01802; CM_pi-yst; 1.
KW   Isomerase.
SQ   SEQUENCE   280 AA;   689D49312CD292D9 CRC64;

Query Match     100.0%;  Score 1434;  DB 2;  Length 280;
Best Local Similarity   100.0%;
Matches  280;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1  MDFMKPETVLDLGNIRDALVRMEDTIIFNFIERSQFYASPSVYKVNQPPIPNFDGSFLDW  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MDFMKPETVLDLGNIRDALVRMEDTIIFNFIERSQFYASPSVYKVNQPPIPNFDGSFLDW  60

Qy   61  LLSQHERIHSQVRRYDALPDEVPFFPNVLEKTFLPKINYPSVLASYADEINVNKEILKIYT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  LLSQHERIHSQVRRYDALPDEVPFFPNVLEKTFLPKINYPSVLASYADEINVNKEILKIYT 120

Qy  121  SEIVPGIAAGSGEQEDNLGSCAMADIECLQSLSRRIHFGRFVABAKFISEGDKIVDLIKK 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  SEIVPGIAAGSGEQEDNLGSCAMADIECLQSLSRRIHFGRFVABAKFISEGDKIVDLIKK 180

Qy  181  RDVEGIEALITNABVEKRILDRLLEKGRAYGTDPTLKFTQHIQSKVKPEVIVKIYQDFVI 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  RDVEGIEALITNABVEKRILDRLLEKGRAYGTDPTLKFTQHIQSKVKPEVIVKIYQDFVI 240

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 10:21:29 ; Search time 4655 Seconds
(without alignments)
2914.599 Million cell updates/sec

Title: US-10-042-059B-2
Perfect score: 1434
Sequence: 1 MDFMKPETVLDLGNIRDALV..........DDAIQKSGGYVDRFLSSGLY 280

Scoring table: BLOSUM62

| | Xgapop | | Xgapext | |
|---|---|---|---|---|
| | 10.0 , | | 0.5 | |
| Ygapop | 10.0 , | Ygapext | 0.5 | |
| Fgapop | 6.0 , | Fgapext | 7.0 | |
| Delop | 6.0 , | Delext | 7.0 | |

Searched: 4708233 seqs, 2422760795 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US10042059/runat_30032005_103454_9002/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -NORM=ext -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10042059 @CGN 1.1.4200 @runat_30032005_103454_9002 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7 -YGAPOP=10

Database : GenEmbl:*

```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1434	100.0	843	6	BD266599	BD266599 Nucleic a
2	1434	100.0	843	6	AX043758	AX043758 Sequence
3	1434	100.0	1652	8	AF204738	AF204738 Pichia an
4	1434	100.0	1655	6	BD266600	BD266600 Nucleic a

5	1434	100.0	1655	6	AX043760	AX043760 Sequence
6	921	64.2	948	6	AR548979	AR548979 Sequence
7	907	63.2	1110000	8	CR382138_20	Continuation (21 o
8	808	56.3	1110000	8	CR382131_20	Continuation (21 o
9	776.5	54.1	771	8	AY693179	AY693179 Saccharom
10	776.5	54.1	2059	8	M24517	M24517 Saccharomyc
11	776.5	54.1	43776	8	YSCARO7A	YSCARO7A Saccharom
12	776.5	54.1	165535	8	SC9499X	Z49219 S.cerevisia
13	752.5	52.5	1110000	8	SCCHRXVI	Z71255 S.cerevisia
14	719	50.1	1110000	8	CR382122_07	Continuation (12 o
15	702.5	49.0	1110000	8	CR380957_11	Continuation (12 o
16	540.5	37.7	869	8	AE016817_01	Continuation (6 of
17	540.5	37.7	869	8	BD178335	BD178335 Transform
18	523	36.5	37301	8	AB116236	AB116236 Rosellini
19	511.5	35.7	1110000	2	SPAC16E8	Z98529 S.pombe chr
20	508.5	35.5	1231	8	AP006499_05	Continuation (6 of
21	508.5	35.5	1250	8	AR404687	AR404687 Sequence
22	506	35.3	3167	8	AK068983	AK068983 Oryza sat
23	481.5	33.6	951	8	AF133241	AF133241 Emericell
24	481.5	33.6	1143	8	BT005306	BT005306 Arabidops
25	481.5	33.6	1223	8	AK117860	AK117860 Arabidops
26	475	33.1	987	8	AR404686	AR404686 Sequence
27	470	32.8	1348	8	AF394889	AF394889 Filobasid
28	466	32.5	1005	8	AK105512	AK105512 Oryza sat
29	465	32.4	1000	8	AX506042	AX506042 Sequence
30	464.5	32.4	1217	6	L47356	L47356 Lycopersico
31	464.5	32.4	1217	6	AR236635	AR236635 Sequence
32	464.5	32.4	1217	6	AR236641	AR236641 Sequence
33	464.5	32.4	1217	8	AX645576	AX645576 Sequence
34	464.5	32.4	1217	8	AX464582	AX464582 Sequence
35	464.5	32.4	1217	8	AF131219	AF131219 Arabidops
36	464	32.4	1207	6	AR236633	AR236633 Sequence
37	464	32.4	1207	6	AR236639	AR236639 Sequence
38	464	32.4	1207	6	AX464574	AX464574 Sequence
39	464	32.4	1207	8	AX464580	AX464580 Sequence
40	451	31.5	1274	8	ATCHMUT	ATCHMUT A.thaliana
41	423	29.5	1020	8	AY089156	AY089156 Arabidops
42	408.5	28.5	1435	8	AR404688	AR404688 Arabidops
43	399	27.8	798	8	AK101220	AK101220 Oryza sat
44	399	27.8	829	8	AX506277	AX506277 Sequence
45	399	27.8	993	6	AX343935	AX343935 Sequence

ALIGNMENTS

RESULT 1
BD266599
LOCUS   BD266599   843 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION   Nucleic acid molecule, containing a nucleic acid which codes for a
          polypeptide with chorismate mutase activity.
ACCESSION   BD266599
VERSION   BD266599.1   GI:33076367
KEYWORDS   JP 2002542788-A/1.
SOURCE   Pichia angusta
  ORGANISM   Pichia angusta
          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
          Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1 (bases 1 to 843)
  AUTHORS   Gellissen,G., Braus,G., Pries,R., Krappmann,S. and Strasser,A.W.
  TITLE   Nucleic acid molecule, containing a nucleic acid which codes for a
          polypeptide with chorismate mutase activity
  JOURNAL   Patent: JP 2002542788-A 1 17-DEC-2002;
          RHEIN BIOTECH GESELLSCHAFT FUER NEUE BIOTECHNOLOGISCHE PROZESSE UND
          PRODUKTE MBH
COMMENT   OS Hansenula polymorpha (yeast)
          PN JP 2002542788-A/1
          PD 17-DEC-2002
          PF 27-APR-2000 JP 2000614405
          PR 27-APR-1999 DB 199 19 124.7
          PI GERD GELLISSEN,GERHARD BRAUS,RALPH PRIES,SVEN KRAPPMANN, PI
          ALEXANDER W STRASSER
          PC C12N15/09,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N15/10 PC
          ,C12N9/90,C12N15/00,

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 10:21:01 ; Search time 584 Seconds
(without alignments)
2838.232 Million cell updates/sec

Title: US-10-042-059B-2
Perfect score: 1434
Sequence: 1 MDFNKPETVLDIGNIRDALV..........DDATQKSGGYVDRFLSSGLY 280

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEl=frame+_p2n:model1 -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10042059/runat_30032005_103453_8996/app_query.fasta_1.455
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DCALIGN=200 -THR_SCORE=pct -HEAPSIZE=500 -MINLEN=0 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10042059 @CGN_1_1_644 @runat_30032005_103453_8996 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY_NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

c		Score	%		DB	ID	Description
c	6	517.5	36.1	989	13	ADR84923	Adr84923 Aspergill
c	7	517.5	36.1	2920	8	ABT17891	Abt17891 Aspergill
c	8	517.5	36.1	2987	8	ABT19705	Abt19705 Aspergill
	9	517.5	36.1	6989	13	ADR84336	Adr84336 Aspergill
	10	508.5	35.5	1231	10	ABX93057	Abx93057 Rice chor
	11	508.5	35.5	1231	12	ADJ555263	Adj555263 Rice chor
	12	481.5	33.6	1223	12	ABX93056	Abx93056 Corn chor
	13	481.5	33.6	1223	12	ADJ555261	Adj555261 Corn chor
	14	470.5	32.8	658	3	AAF14469	Aaf14469 Aspergill
	15	466	32.4	1005	6	ABZ12932	Abz12932 Arabidops
	16	464.5	32.4	1217	6	ABA91395	Aba91395 Arabidops
	17	464	32.4	1217	6	ABA91401	Aba91401 Arabidops
	18	464	32.4	1207	6	ABA91393	Aba91393 Arabidops
	19	464	32.4	1207	6	ABA91399	Aba91399 Arabidops
	20	451	31.5	1274	3	AAC33324	Aac33324 Arabidops
c	21	423	29.5	1020	10	ABX93058	Abx93058 Soybean c
c	22	423	29.5	1020	12	ADJ555265	Adj555265 Soybean c
	23	399	27.8	798	6	ABZ11167	Abz11167 Arabidops
	24	399	27.8	993	6	ABA99614	Aba99614 Construct
	25	399	27.8	1006	6	ABA99613	Aba99613 A. thalia
	26	399	27.8	1006	6	ABA91400	Aba91400 Arabidops
c	27	399	27.8	1006	6	ABA91394	Aba91394 Arabidops
	28	355	24.8	622	13	ACN50132	Acn50132 Cotton no
	29	346	24.1	580	13	ACN57510	Acn57510 Cotton gy
	30	325	22.7	864	12	AD744958	Ad744958 Plant cDN
	31	313	21.8	780	12	ABX93059	Abx93059 Wheat cho
	32	313	21.8	780	12	ADJ555267	Adj555267 Wheat cho
	33	309.5	21.6	5099	6	AAD29643	Aad29643 Arabidops
	34	309.5	21.6	5099	6	AAD29643	Aad29643 Arabidops
	35	286	19.9	5099	13	ABA91396	Aba91396 Arabidops
	36	269	18.8	524	13	ADR64235	Adr64235 Cotton CD
c	37	245	17.1	600	13	ADR64815	Adr64815 Cotton CD
	38	240	16.7	542	13	ACN57601	Acn57601 Cotton gy
	39	240	16.7	579	10	ABX93054	Abx93054 Wheat cho
	40	213.5	14.9	579	13	ACN50205	Acn50205 Cotton no
c	41	211.5	14.7	584	13	ACN50257	Acn50257 Cotton no
	42	211.5	14.7	525	10	ABX93053	Abx93053 Soybean c
	43	209	14.6	258	12	ADJ555255	Adj555255 Soybean c
	44	195.5	13.6	5176	6	ABL71526	Abl71526 Corn tass
c	45	191	13.3	507	3	ABA91397	Aba91397 Arabidops
						AAF07985	Aaf07985 Fusarium

RESULT 1
AAC81949
ID AAC81949 standard; DNA; 843 BP.
XX
AC AAC81949;
XX
DT 15-SEP-2003 (revised)
DT 01-MAR-2001 (first entry)
XX
DE H. polymorpha chorismate mutase DNA.
XX
KW Chorismate mutase; prephenate; selection marker; auxotrophic yeast; de.
XX
OS Pichia angusta.
XX
PN WO20006507l-A1.
XX
PD 02-NOV-2000.
XX
PF 27-APR-2000; 2000WO-EP003844.
PR 27-APR-1999; 99DE-01019124.
XX
PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
PI Gellissen G, Braus G, Pries R, Krappmann S, Strasser AW;
XX
DR WPI; 2000-687355/67.
DR P-PSDB; AAB11451.

ALIGNMENTS

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1434	100.0	843	3	AAC81949	Aac81949 H. polymo
2	1434	100.0	1655	3	AAC81950	Aac81950 H. polymo
3	921	64.2	801	12	ADP98547	Adp98547 C. albica
4	587.5	41.0	801	13	ADR85510	Adr85510 Aspergill
5	540.5	37.7	869	8	AAL50202	Aal50202 M sterili

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 10:42:01 ; Search time 195 Seconds
                                      (without alignments)
                                      2349.526 Million cell updates/sec

Title: US-10-042-059B-2
Perfect score: 1434
Sequence: 1 MDFMKPETVLDIGNIRDALV.........DDATQKSGGYVDRFLSSGLY 280

Scoring table: BLOSUM62
                 Xgapop 10.0   Xgapext   0.5
                 Ygapop 10.0   Ygapext   0.5
                 Fgapop   6.0   Fgapext   7.0
                 Delop   6.0   Delext   7.0

Searched: 1 MDFMKPETVLDIGNIRDALV.........DDATQKSGGYVDRFLSSGLY 280

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Searched: 1202784 seqs, 818183359 residues

Command line parameters:
-MODEL=frame+_p2n:model
-Q=/cgn2.1/USPTO_spool/US10042059/runat.30032005_103456_9019/app_query.fasta_1.455
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0
-Issued_Patents_NA -QFMT=fastap -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -UNITS=bits -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10042059 @CGN_1_1_177 @runat.30032005_103456_9019 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
            1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Post-processing: Minimum Match 0%

Pred: No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	921	64.2	948	4	US-09-248-796A-4110	Sequence 4110, Ap
2	508.5	35.5	1231	4	US-09-454-279-13	Sequence 13, Appl
3	481.5	33.6	1223	4	US-09-454-279-11	Sequence 11, Appl
4	464.5	32.4	1217	4	US-09-610-040-4	Sequence 4, Appl
5	464.5	32.4	1217	3	US-10-267-763-4	Sequence 4, Appl
6	464.5	32.4	1217	4	US-09-610-040-10	Sequence 10, Appl
7	464.5	32.4	1217	4	US-10-267-763-4	Sequence 4, Appl
8	464	32.4	1207	3	US-09-610-040-2	Sequence 2, Appl
9	464	32.4	1207	3	US-10-267-763-10	Sequence 10, Appl
10	464	32.4	1207	4	US-09-610-040-8	Sequence 8, Appl
11	464	32.4	1207	4	US-10-267-763-2	Sequence 2, Appl
12	423	29.5	1020	4	US-10-267-763-8	Sequence 8, Appl
					US-09-454-279-15	Sequence 15, Appl

13	399	27.8	1006	3	US-09-610-040-3	Sequence 3, Appl
14	399	27.8	1006	3	US-09-610-040-9	Sequence 9, Appl
15	399	27.8	1006	4	US-10-267-763-3	Sequence 3, Appl
16	399	27.8	1006	4	US-10-267-763-9	Sequence 9, Appl
17	313	21.8	780	3	US-09-454-279-17	Sequence 17, Appl
18	309.5	21.6	5099	3	US-09-610-040-5	Sequence 5, Appl
19	309.5	21.6	5099	4	US-10-267-763-5	Sequence 5, Appl
20	240	16.7	579	4	US-09-454-279-7	Sequence 7, Appl
21	211.5	14.7	525	3	US-09-454-279-5	Sequence 5, Appl
22	195.5	13.6	5176	3	US-09-610-040-6	Sequence 6, Appl
23	195.5	13.6	5176	4	US-10-267-763-6	Sequence 6, Appl
24	126	8.8	601	4	US-09-454-279-3	Sequence 3, Appl
25	120.5	8.4	552	4	US-09-454-279-1	Sequence 1, Appl
26	99	6.9	1664976	4	US-08-916-421B-1	Sequence 1, Appl
27	99	6.9	1664976	4	US-09-692-570-1	Sequence 1, Appl
28	94	6.6	1105	4	US-09-636-077A-17	Sequence 17, Appl
29	94	6.6	1105	4	US-09-636-060C-17	Sequence 17, Appl
30	94	6.6	1105	4	US-09-998-552-17	Sequence 17, Appl
31	94	6.6	1105	4	US-09-248-796A-2026	Sequence 2026, Ap
32	94	6.6	1105	4	US-09-636-596C-17	Sequence 17, Appl
33	94	6.6	1105	3	US-10-023-894-15	Sequence 15, Appl
34	94	6.6	1105	3	US-10-306-686-17	Sequence 17, Appl
35	94	6.6	1232	4	US-09-270-767-10717	Sequence 10717, A
36	93.5	6.5	59828	4	US-09-949-016-16238	Sequence 16238, A
37	92	6.4	2154	4	US-09-583-110-1968	Sequence 1968, Ap
38	92	6.4	2166	4	US-09-107-433-216	Sequence 216, Appl
39	90	6.3	879	4	US-09-107-532A-87	Sequence 87, Appl
40	89.5	6.2	1923	4	US-09-248-796A-2026	Sequence 2026, Ap
41	89.5	6.2	2441	4	US-09-051-465-1	Sequence 1, Appl
42	89.5	6.2	2871	3	US-09-051-465-2	Sequence 2, Appl
43	89.5	6.2	2871	3	US-09-051-465-4	Sequence 4, Appl
44	89	6.2	1005	4	US-09-934-901-5	Sequence 5, Appl
45	89	6.2	1005	4	US-09-934-868-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-248-796A-4110
; Sequence 4110, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4110
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4110

Alignment Scores:
Pred. No.:   8.57e-116
Score: 921.00
Percent Similarity: 82.63%
Best Local Similarity: 65.64%
Query Match: 64.23%

	Length:	948
	Matches:	170
	Conservative:	44
	Mismatches:	45
	Indels:	0
	Gaps:	0

US-10-042-059b-2 (1-280) x US-09-248-796A-4110 (1-948)

Qy   1 MetAspPheMetLysProGluThrValLeuAspLeuGlyAsnIleArgAspAlaLeuVal 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 142 ATGGATTTTATGAAACCAGAAACTGTGCTTGATCTTGGCAACATCCGTCAAGCATTGGTA 201

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 11:54:44 ; Search time 622 Seconds
(without alignments)
2724.594 Million cell updates/sec

Title: US-10-042-059B-2
Perfect score: 1434
Sequence: 1 MDFMKPBTVLDLGNIRDAIV.........DDATQKSGGYVDRFLSSGLY 280

Scoring table: BLOSUM62
Xgapop 10.0   Xgapext 0.5
Ygapop 10.0   Ygapext 0.5
Fgapop 6.0   Fgapext 7.0
Delop 6.0   Delext 7.0

Searched: 5607317 seqs, 302624599 residues

Total number of hits satisfying chosen parameters: 11214634

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10042059/runat_30032005_103458_9071/app_query.fasta_1.455
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10042059 @CGN_1_1_688 @runat_30032005_103458_9071
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match *	Length DB	ID	Description
	1	1434	100.0	843	13	US-10-042-059A-1	Sequence 1, Appli
	2	1434	100.0	1655	13	US-10-042-059A-3	Sequence 3, Appli
	3	.921	64.2	807	19	US-10-741-849-6032	Sequence 6032, Ap
	4	540.5	37.7	869	18	US-10-472-587-26	Sequence 26, Appl
	5	517.5	36.1	2920	15	US-10-128-714-249	Sequence 249, App
	6	517.5	35.5	2987	15	US-10-128-714-5249	Sequence 5249, Ap
c	7	508.5	35.5	1016	17	US-10-437-963-93956	Sequence 93956, A
	8	508.5	35.5	1231	13	US-09-454-279-13	Sequence 13, Appl
	9	508.5	34.9	1231	17	US-10-739-930-4455	Sequence 4455, Ap
	10	500.5	34.7	1548	17	US-10-425-1114-7285	Sequence 7285, Ap
	11	497.5	34.7	1146	17	US-10-425-1114-7285	Sequence 7285, Ap
	12	495.5	34.6	1423	17	US-10-424-599-123820	Sequence 123820, A
	13	489.5	34.1	1376	17	US-10-425-1115-102347	Sequence 102347, A
	14	486	33.9	1500	18	US-10-767-701-14353	Sequence 14353, A
	15	486	33.6	1223	9	US-09-454-279-11	Sequence 11, Appl
	16	481.5	33.6	1223	17	US-10-624-061-11	Sequence 11, Appl
	17	481.5	33.4	1430	18	US-10-739-930-4454	Sequence 4454, Ap
	18	479.5	33.4	1397	18	US-10-425-115-102346	Sequence 102346, A
	19	479.5	33.4	658	18	US-10-65-047-6992	Sequence 6992, Ap
	20	478.5	32.8	1368	18	US-10-425-115-102349	Sequence 102349, A
	21	478.5	32.8	1005	9	US-09-938-842A-737	Sequence 737, App
	22	469.5	32.7	1005	18	US-09-938-842A-737	Sequence 737, App
	23	469.5	32.5	1217	11	US-10-267-763-4	Sequence 4, Appli
	24	466	32.5	1217	14	US-10-267-763-2	Sequence 2, Appli
c	25	466	32.4	1207	14	US-10-267-763-10	Sequence 10, Appli
c	26	464.5	32.4	1207	14	US-10-437-963-49372	Sequence 49372, A
	27	464.5	32.4	1299	18	US-09-938-842A-972	Sequence 972, App
	28	464.5	32.4	1020	9	US-09-938-842A-972	Sequence 972, App
	29	464	31.5	1020	18	US-10-267-763-8	Sequence 8, Appli
	30	464	31.5	1079	17	US-10-739-930-621	Sequence 621, Appl
	31	451	29.5	1653	17	US-09-454-279-15	Sequence 15, Appl
c	32	451	29.5	768	18	US-10-624-061-15	Sequence 15, Appl
	33	423	29.5	798	9	US-10-425-114-7639	Sequence 7639, Ap
c	34	423	29.2	798	18	US-10-425-599-93036	Sequence 93036, A
c	35	423	27.8	1006	14	US-10-437-963-49372	Sequence 49372, A
	36	419	27.8	1006	14	US-10-267-763-3	Sequence 3, Appli
	37	399	27.8	1010	14	US-10-267-763-9	Sequence 9, Appli
	38	399	27.8	1006	18	US-10-267-763-9	Sequence 9, Appli
	39	399	26.8	1314	17	US-10-425-115-102351	Sequence 102351, A
	40	385	26.6	1066	18	US-10-425-115-151470	Sequence 151470, A
	41	385	26.1	1083	17	US-10-425-115-151470	Sequence 151470, A
	42	381	26.1	1197	18	US-10-425-114-1332	Sequence 1332, App
c	43	374.5	24.8	921	18	US-10-425-115-115559	Sequence 115559, A
c	44	373	24.8	622	18	US-10-437-963-83564	Sequence 83564, A
c	45	356	24.8	2154	18	US-10-437-963-83557	Sequence 83557, A
		355	24.5	2154	18	US-10-437-963-83557	Sequence 83557, A
		346	24.1	580	18	US-10-021-323-12291	Sequence 12291, A

RESULT 1
US-10-042-059A-1

; Sequence 1, Application US/10042059A
; Publication No. US20020197704A1
; GENERAL INFORMATION:
; APPLICANT: Rhein Biotech Gesellschaft fur neue biotechnologische Prozesse und
; APPLICANT: Produkte mbH
; TITLE OF INVENTION: Nucleic Acid Molecule Containing a Nucleic Acid Coding for a Pol
; TITLE OF INVENTION: with Chorismate Mutase Activity
; FILE REFERENCE: PA1055BUS-019
; CURRENT APPLICATION NUMBER: US/10/042,059A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: DE 199 19 124.7
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Hansenula polymorpha

ALIGNMENTS

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: March 31, 2005, 10:22:54 ; Search time 7476 Seconds
(without alignments)
1425.628 Million cell updates/sec

Title: US-10-042-059B-2
Perfect score: 1434
Sequence: 1 MDFMKPETVLDLGNIRDALV...........DDATQKSGGYVDRFLSSGLY 280

Scoring table:
BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10042059//runat_30032005_103455_9011/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10042059 @CGN_1_1_4352 @runat_30032005_103455_9011 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*

1:   gb_est1:*
2:   gb_est2:*
3:   gb_htc:*
4:   gb_est3:*
5:   gb_est4:*
6:   gb_est5:*
7:   gb_est6:*
8:   gb_gss:*
9:   gb_gss1:*
9:   gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	758	52.9	1185	9	CNS06T5P	AL414131 T7 end of
	2	570	39.7	948	7	CO005472	C0005472 EST793807
c	3	506.5	35.3	574	8	BZ303243	BZ303243 KD2284.p1
	4	481.5	33.6	1222	7	AY103806	AY103806 Zea mays
c	5	472	32.9	788	7	CO165801	CO165801 FLD1_57_E
c	6	470	32.8	1329	3	CNS0A727	BX822129 Arabidops
c	7	469	32.7	937	9	CNS06TV7	BX822129 Arabidops
c	8	469	32.7	1052	9	CNS06XEJ	AL415049 T3 end of
c	9	465	32.4	1265	3	CNS0A6WL	BX824460 Arabidops

	Result No.	Score	%	DB	ID	Description
	10	460.5	32.1	7	CF475712	CF475712 RTWW2_11
	11	458.5	32.0	7	CO120964	CO120964 GR_Eb024
	12	448	31.2	7	CO005473	C0005473 EST793808
	13	442.5	30.9	7	CB619882	CB619882 OSIIBa04L
	14	436.5	30.4	6	CD423256	CD423256 SA1_28_B1
	15	436.5	30.4	5	BX253798	BX253798
	16	434.5	30.3	7	CK196770	CK196770 tric075xb
	17	434	30.3	7	CF877247	CF877247 tric075xb
	18	434	30.1	6	CB905545	CB905545 tric075xb
c	19	418	29.1	6	AG273371	AG273371 Cyanidios
	20	416.5	29.0	1	AJ612241	AJ612241
	21	415.5	29.0	5	BQ989273	BQ989273 QGF17D13.
	22	413.5	28.8	8	CNS23137	CNS23137 GQ015M01.
c	23	409	28.8	6	CA080410	CA080410 SCVPAM05.
	24	407	28.4	6	CB619881	CB619881 OSIIBa04L
	25	404	28.2	1	AJ833468	AJ833468 SA1_28_B1
	26	403.5	28.1	6	CD423358	CD423358 Hyacinthu
c	27	401.5	28.1	3	AY389761	AY389761 OG04A12T3
	28	400.5	27.9	6	BF430733	BF430733 OG04A12T3
	29	398.5	27.8	6	CN149106	CN149106 WOUND1_60
	30	394.5	27.5	7	CO366784	CO366784 RTKI_30_C
	31	392	27.3	6	CO716620	CO716620 GR_Ea30A
	32	392	27.3	6	CO071737	CO071737 GR_Ea30A
c	33	387.5	27.0	6	CA293493	CA293493 SCSGLV100
	34	387	27.0	6	CO071299	CO071299 GR_Ea29C
c	35	384.5	26.8	6	CB622596	CB622596 OSIIEa09I
	36	383.5	26.8	6	CF444592	CF444592 EST680937
	37	383.5	26.7	6	CF635400	CF635400 zmrww00_0
	38	382	26.6	6	CD982471	CD982471 QAK504.x
c	39	381	26.6	7	CO165812	CO165812 FLD1_57_F
c	40	380.5	26.5	7	CO427426	CD427426 SA1_30_H0
	41	379.5	26.5	7	CV274873	CV274873 WS0174.B2
c	42	378	26.4	6	CA121888	CA121888 SCJFLR105
	43	376.5	26.3	6	BS577489	BS577489 L48-2335lr
	44	374	26.1	2	BE494722	BE494722 WHE1257_A
	45	374	26.1	7	CO094261	CO094261 GR_Ea16E

ALIGNMENTS

RESULT 1
CNS06T5P

LOCUS   CNS06T5P   1185 bp   DNA   linear   GSS   05-JUL-2001
DEFINITION   Yarrowia lipolytica, genomic survey sequence.
ACCESSION   AL414131
VERSION   AL414131.1   GI:12186949
KEYWORDS   GSS.
SOURCE   Yarrowia lipolytica
  ORGANISM   Yarrowia lipolytica
           Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
           Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1   (bases 1 to 1185)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
           Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
           de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
           Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
           Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
           Wincker,P. and Weissenbach,J.
  TITLE   Genomic exploration of the hemiascomycetous yeasts: 1. A set of
           yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2   (bases 1 to 1185)
  AUTHORS   Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Peynerol,C.,
           Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE   Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
           lipolytica
  JOURNAL   FEBS Lett. 487 (1), 95-100 (2000)
  MEDLINE   20584727
   PUBMED   11152892
REFERENCE   3   (bases 1 to 1185)